A TEXT BOOK ON LIGHT

By the Same Author

A TEXT BOOK ON HEAT. 9*s*.

A TEXT BOOK ON LIGHT

4389 Cloth

BY

A. W. BARTON, M.A., PH.D.

Headmaster of King Edward VII School, Sheffield
Sometime Senior Scholar of Trinity College, Cambridge,
and Assistant Master at Repton School
Author of "A Text Book on Heat"

With Halftone Plates
and Diagrams in the Text

LONGMANS, GREEN & CO. LTD.
LONDON • NEW YORK • TORONTO
BOMBAY • CALCUTTA • MADRAS

LONGMANS, GREEN AND CO. LTD.
OF PATERNOSTER ROW
43 ALBERT DRIVE, LONDON, S. W. 19
NICOL ROAD, BALLARD ESTATE, BOMBAY
17 CHITTARANJAN AVENUE, CALCUTTA
36A MOUNT ROAD, MADRAS

LONGMANS, GREEN AND CO.
55 FIFTH AVENUE, NEW YORK

LONGMANS, GREEN AND CO.
215 VICTORIA STREET, TORONTO

First Published 1939
Reprinted 1943

CODE NUMBER : 85317

Printed in India at the Diocesan Press, Madras

PREFACE

THIS book has been written for students with an elementary knowledge of Light, who are reading for University Scholarships, the various Higher Certificate Examinations, and University Intermediate Examinations. It is hoped that it will also be adequate for those readers who do not wish to continue the study of the subject beyond the Pass Degree standard, provided it is supplemented with suitable lecture notes. The convention of signs adopted is the first of the two recommended in the recent report of the Physical Society.

The teaching of Science specialists at Schools and Universities has been practised for some fifty years, and well-tried methods have become established. The subject matter is presented logically and as simply as truth will allow, only finally accepted theories being expounded. This method has the virtues of simplicity and definiteness and the doubtful virtue of leading to success in examinations, for, if the student fails to understand the ideas presented to him, he has only to learn them by heart to gain good marks. But it has some disadvantages which are serious in the case of students at the less advanced stages. The work may become somewhat dull, because only one theory is presented, and, since ideas are put forward as if they were inevitably true, there is little opportunity for discussion and the exercise of the critical faculty. When the student trained in this way begins to do research, he is often bewildered because he is faced with a type of situation which is quite strange to him.

A different method of presentation has been adopted in this book; it is not a logical method, nor is it historical in the sense that there is any reference to the life history of pioneer workers. But it is historical in the sense that the discovery of facts and the development of ideas is presented as they occurred historically. When the set of facts which were historically known at a certain date have been discussed and established, the theories which were put forward to correlate them are expounded solely in the light of those facts. There is no dogmatic suppression of one theory because the teacher, with his wider knowledge of facts, is aware that later evidence will rule it out. The reader is therefore treated as if he were at the frontiers of knowledge; he is made to exercise his critical faculty in deciding which of the two theories is the more satisfactory; his interest is aroused by the uncertainty of the situation and the conflict of ideas, and he sees how new lines of investigation are suggested by the deductions made from the theories. So he learns that

scientific knowledge grows by a judicious interplay between experiment and theory, the theory correlating experimental facts and suggesting further experiments. In this way the mass of evidence grows until one theory becomes untenable and yet another advance has been attempted and made good. It is felt that students trained in this way will be better able to do original work and to solve problems when they commence their professional careers.

This method of teaching will be called the dynamic method, because it tries to trace the growth of the ideas correlating the experimental evidence and the modification of those ideas by the further investigations which they have suggested. It has been developed in the case of Light in the following way. Lenses and mirrors occur naturally and their power to produce images was known to the Greeks and Romans. Geometrical optics is developed as an attempt to analyse this power of lenses and mirrors to produce such images, and it is shown how this analysis led to an extension of their uses and to an improvement in the definition of the images. Later the discovery of the telescope, which was probably accidental, led to the extension of this analysis to combinations of lenses and mirrors and opened up the new field of optical instruments, which has revealed to mankind both the depths of space and the infinitesimal detail of the structure of living beings and inanimate matter. Finally, the recent advances in photometric technique are shown to be due to the growth of industrial civilisation, with its large urban populations. As the simple fundamental facts about Light have now been established, the theories to explain them are discussed. The attempts of both the corpuscular theory and the wave theory are dealt with, and it is interesting to observe that the explanations of rectilinear propagation and reflection on the corpuscular theory always make a strong appeal to the student and seem to him to be superior to those of the wave theory. This illustrates another advantage of the dynamic method of teaching, in that its presentation of the subject matter is likely to be in harmony with the student's mental development. It is seen that neither theory is completely satisfactory, the usual situation when only a limited range of facts is known. The wave theory ultimately triumphs, not because of any one crucial experiment, but because it suggests the search for interference and diffraction, both of which are finally discovered and lead to ways of measuring the wave-length of light. The book concludes with an elementary discussion of the photo-electric effect and Compton effect, which show in a simple way the necessity for the revival of the corpuscular theory in a new form, and so, as with Newton's Optics, this volume closes with some " Queries ".

It must be left for others to decide if there is any virtue in this type of presentation, but it is hoped that the reader will acquire a keen critical faculty and the power to solve problems, which are so important in a changing world ; and that he will learn that not only does new knowledge

revolutionise old trades, but that urgent practical problems also lead to new knowledge which is of interest for its own sake ; and lastly, as a result of the attempt made to describe the subject matter against its relevant social background, it is hoped that the reader will become conscious of the social responsibility of scientists.

A selection of examples is given at the end of each chapter, some of which have been composed by the author and the remainder taken from recent examination papers by kind permission of the Cambridge University Press, the Northern Universities Joint Matriculation Board, the University of London, and the Examining Boards to the two principal Groups of Oxford Colleges. The source of each of these examples is given in brackets after the question.

I must acknowledge my indebtedness to the standard works on Light by Newton and Mach and also to Martin's "Introduction to Applied Optics". I should also like to express my sincere thanks to my friend, Dr. A. Wood, University Lecturer in Experimental Physics at the University of Cambridge, who has read through the manuscript and made a number of valuable suggestions. I must also thank my colleague, Dr. J. W. Mitchell, of Repton School, who gave me the benefit of his great experience in spectroscopy in writing Chapter 16, and who has worked out the answers to the examples. Finally, I am obliged to Mr. J. W. Cottingham, of Barnsley Grammar School, and to Dr. J. W. Mitchell for the beautiful photographs from which the Plates have been prepared.

KING EDWARD VII SCHOOL,
SHEFFIELD.
June, 1939.

The sources of the examples are indicated as follows:

Camb. Schol.	Entrance Scholarships at Cambridge Colleges.
Oxford Schol.	Entrance Scholarships at Oxford Colleges.
O. and C.	Oxford and Cambridge Schools Examination Board Higher Certificate.
N.U.J.B.	Northern Universities Joint Matriculation Board Higher Certificate.
London	London University Higher School Certificate.
Lond. B.Sc.	London University B.Sc. General Honours Examination.
Tripos, Part I.	Part I of the Natural Sciences Tripos, Cambridge University.

CONTENTS

LIST OF PLATES

Chapter I

INTRODUCTORY

1. FUNDAMENTAL IDEAS

The foundation of the "Royal Society of London for Improving Natural Knowledge" in the year 1660 marks the beginning of serious scientific work in Western civilisation, and science has now reached a vigorous maturity and is playing a big part not only in the control and shaping of man's material environment but also in the revolution of the fundamental ideas in philosophy and religion. When maturity has been reached, it is natural to pause, to look back, and to ask, What is science? What are its aims? How does it achieve them? Perhaps the best answer to the first two questions is that given by Dingle: "Science is the recording, augmentation, and rational correlation of those elements of our experience which are actually or potentially common to all normal persons." We may pause to elaborate this statement a little. Experiences may be taken to include all things of which we are conscious except rational ideas. They include not only sense perceptions, such as warmth, sound, sight, and the countless sensations to which they give rise, but also such experiences as mental states, which are treated scientifically in psychology, and tribal beliefs and customs, which are discussed in anthropology. But science restricts itself to those experiences common to all persons; this expresses what is inherent in the word **fact**. A fact may be defined as an experience which is agreed upon by all normal persons. One or two examples will make this clear. If we look into a pond with a still surface by the side of a tree, we shall see a representation of the tree upside down in the pond. This representation of the tree will correspond in every detail to the actual tree and it will disappear if the surface of the pond is disturbed by waves. This experience is agreed upon by all normal persons; notice that a blind person cannot be regarded as normal, which is quite natural, as he lacks the sense of sight. Everyone will agree about what they see in the representation of the tree down to quite small details. Such an experience is a fact; and science deals only with such facts. It deliberately selects those elements of the immense variety of our experience which do satisfy this test of being common to all normal persons. The visions which some persons claim to have seen in a spiritualistic séance are not facts, since they cannot be experienced by all normal persons; it is only a few persons who have

ever seen them. Again, the experience that water turns anhydrous copper sulphate blue is a fact, because everyone agrees upon it; anyone who tries the experiment gets the same result. A failure can always be explained in a rational way: the white powder was not anhydrous copper sulphate, the liquid was not water, and so on. But the experience that a specially shaped twig moves downwards when it is carried over a place rich in water is not a fact, since very few persons can observe the movement when carrying the twig. It is not denied that a few persons may be able to detect the presence of water below the ground in this way, but it is not a fact, since the experience is not common to all normal persons.

When facts have been recorded, they are augmented. They are **classified**, ideally into a mathematical equation. This enables their significance to be the better appreciated. The immense number of observations about the motion of the planets round the sun were augmented by Kepler into his three laws of planetary motion, which are a neat expression, readily apprehended by the mind, of the motion of the planets and the relations between those motions.

This important stage of science clears the way for the third vital stage, the **rational correlation** of the facts. This is just an explanation of the facts, a **theory** to explain the facts. What does this word "explanation" mean? A simple example will make it clear. When we say that we have explained the motion of the moon round the earth as due to gravitation, we have merely shown that the earth attracts the moon just as it attracts a stone; we have correlated the "falling moon" with the falling stone. We have rendered the rare less mysterious by showing its connection with an everyday occurrence. This misleads some enthusiasts into thinking that we really know *why* the moon goes round the earth. But it did not deceive Newton, who said, "Do not ascribe unto me any *explanation* of the force of gravity." Science states *how* things happen, not *why* they take place. When we say that we have explained the magnetism of the elementary magnets as due to electrons spinning or describing orbits about the nucleus of the atom, we have just correlated this magnetism with that due to an electric current in a coil of macroscopic size. This process of rational correlation of the facts is in a sense the construction of a **working thought model** of the facts. What we experience when we handle a football is the pressure of the air with which it is blown up, and further experience with gases under pressure leads us to augment the facts into the law that pressure and volume are inversely proportional to each other. Our working thought model of the gas in the football is a collection of tiny particles, the molecules of the gas, moving at speeds of about a mile per second and going in straight lines except when they are deflected by encounters with other molecules. The pressure of the gas is due to the rebounding of the molecules off the bladder of the football. By applying the laws of

mechanics to this working thought model, we can deduce an expression for the pressure of the gas and show that the model obeys the above law. We have correlated the gas in the football with the laws of mechanics, which are obeyed by flying bullets, billiard balls, and motor-cars. The theory can be called a working thought model, because it is a representation of the experience, a kind of model of it ; a working model, because it works like a machine ; and a thought model because it only exists in our minds. Perhaps this is its very power, because it can be divested of all those awkward practical irregularities which spoil actual models, and since it exists in our minds we have to deduce what it will do ; it can thus be subjected to that powerful tool, mathematical reasoning.

Such a theory must do one further thing. It must not only correlate the known facts, but it must suggest the existence of new facts ; it should suggest new and fruitful lines of enquiry. The correlation of the facts about planetary motion with the familiar laws of falling bodies at the earth's surface by the law of gravitation had a long sequence of successes until the nineteenth century. It was then found that the motion of the planet Uranus did not quite fit in with the law, but the law suggested that these irregularities were due to the presence of another hitherto undiscovered planet. Moreover, it told the astronomer where to find the planet, and when he looked it was there ! So the theory acts as a signpost in the unknown land, leading us to the hidden treasure. And this brings us to the last characteristic of science : it begins with observation and ends with observation. And its observations are restricted to those common to all normal persons. Thus such experiences as goodness, the beauty of a picture lie outside its ken ; but science does not deny their existence, only its technique cannot deal with them.

What is its technique ? In describing the aim of science, we have dealt with its technique. It can be summed up in four steps : record the facts, the experiences common to all normal persons ; augment or classify them into a law, best of all a mathematical equation ; correlate the facts rationally or invent a theory to explain them ; and that theory or working thought model should lead to a search for new facts. There is just one article of faith in science, a belief in a sequence of natural events, a belief in the existence of law and order in nature. It has now been abundantly justified by experience.

It is the aim of this book to show how the subject of Light has developed along these lines and to indicate at what stage our knowledge has arrived. Some indication will be given of recent additions to knowledge, but a full discussion of them is out of place in a book of this standard, and must be sought in more advanced books. The fundamental notion in the subject of Light is the act of seeing or the sensation of sight. Such a thing cannot be described or explained, it can only be designated ; but the sensation is familiar to all normal persons. We are also familiar with the fact that certain bodies, such as the sun, the stars, a piece of

hot coal, or the hot filament of an electric lamp, can cause this sensation, while most bodies cannot of themselves do so. The former class of bodies are called self-luminous, the remainder being non-luminous. It is also well known that luminous bodies can cause non-luminous bodies in their neighbourhood to become luminous, much as an electric charge can induce charges into a conductor nearby. **Light is the " thing " which a luminous body sends to the eye and which causes the sensation of sight when it strikes the eye.** It should be noticed that the word " thing " is used without committing ourselves in any way as to the nature of light, and similarly the word " send " does not imply that the speed of the process is finite. We must retain an open mind on these points until sufficient evidence is available to settle them. It should also be emphasised that light has no concern with the actual sensation of seeing, nor is it concerned with the way in which the eye sends its message to the brain, such topics being treated under physiology. In the science of Light we restrict ourselves to a consideration of the physical process by which a luminous body produces a physiological sensation. The term " physical " refers here to something occurring apart from the observer, either in space or in some material substance.

2. RECTILINEAR PROPAGATION

The first property of light is that **it travels in straight lines,** a fact so obvious that it was known to the very earliest peoples who investigated nature. Anyone who has observed rays of sunlight through a rift in the clouds or has seen the beam of a motor-car headlight on a misty night can see the truth of this law for himself. It is also verified by the pin-hole camera, which was first made generally known by Porta in the middle of the sixteenth century. He found that if a small hole was made in the shutters of a room, an inverted image of the objects outside in their true colours was thrown on the opposite wall. The eye of the Nautilus is an example of a natural pin-hole camera. This law is also verified by the production of shadows, of which the most striking examples are the eclipses of the sun and moon. But if a shadow is produced in the laboratory, one rather interesting and significant point will be noticed, namely, that the edge of the shadow is *never quite sharp*, however sharp the edge of the object producing it may be. Of course, the edge can be made sharp by putting the screen very close to the object, but this does not give the penumbra due to the finite size of the source a real chance to develop. But if the screen is about the same distance from the object as the object is from the source, the edge of the shadow will never be quite sharp. It may be said that this is due to the fact that it is impossible to produce a strictly point source of light. This may be true; but it may also be true that light does not travel quite in straight lines. It is certain that the deviation from rectilinear propagation must be small, but the evidence does not justify an assertion that such a departure does not

exist. It is interesting to notice, in passing, that similar uncertainties exist in many scientific laws, which must always be accepted with due regard for experimental error, and which are therefore always liable to revision in the light of more accurate experiments. With this reservation, then, we may regard rectilinear propagation as the fundamental law of light, and this leads to the idea of a **ray** of light, which is **the straight line between two points in an isotropic medium along which light is propagated.** An isotropic medium is one whose properties are the same in all directions, and this restriction is necessary as rectilinear propagation is only true in such media. It will be seen that a ray of light is a somewhat abstract conception ; it really comes to this : we artificially split the beam of light emerging from a searchlight, for example, into an infinite number of rays of light ; we do not know yet whether the rays have any real physical existence, but they are a convenient concept which enables us to analyse what happens, and the results obtained do fit many facts, so we shall retain the conception until we find it is at variance with the evidence.

3. REFLECTION AND REFRACTION

The phenomenon of reflection, which refers to the fact that when light strikes a new medium some of it is thrown back into the original medium, must have been known from the earliest times, since perfect reflections of buildings, trees, and other objects can be seen in still water. It is not surprising to find, therefore, that not only did the Greeks know the law of rectilinear propagation, but that they also discovered the second of the two laws of reflection of light, which state that **the reflected ray lies in the same plane as the incident ray and the normal to the reflecting surface, and that the two rays make equal angles with the normal.** It is hardly necessary to say that this law can only be established by experimental investigation, and it is quite simply done by sending a ray of sunlight, or one produced artificially, on to a mirror and measuring the angles of incidence and reflection. Full details can be found in any elementary text-book on Light. It follows from these laws that all the rays of light coming from a point source which strike a plane mirror appear to come after reflection from a point as far behind the mirror as the source is in front. In other words, the mirror forms a reflection of the point source behind the mirror. We can regard a finite object as made up of a large number of point sources, so the mirror will form a representation of the object as far behind the mirror as the object is in front. Such a representation of an object produced by an optical arrangement is called an image of the object. We have already seen how a pin-hole camera produces an image of objects, but there is a difference between the image it produces and that formed by a plane mirror. The former can be cast on a screen ; the rays of light from a point on the object actually converge to the corresponding point on the image, which is called a real image. But the image produced by a plane

mirror is called a virtual image, because it cannot be cast on to a screen and the rays of light from a point on the object only appear to diverge from the corresponding point on the image. We shall come across many other cases of real and virtual images. This analysis of how a plane mirror produces images can also be extended to show why lateral inversion occurs in plane mirrors, and to the effects when a mirror is rotated and several mirrors are used to produce images. The reader must consult an elementary book on the subject for full details of these points.

We have seen that when light strikes a new medium, some of it is reflected but some goes on into the new medium. Euclid was the first to observe some effects of the fact that the ray of light bends towards the normal when it passes from a medium such as air to a denser medium such as water or glass, although he did not himself discover the fact. It is interesting to notice that a further 2000 years elapsed before the law of refraction was discovered by Snell. An Egyptian astronomer, Ptolemy, who lived in the second century A.D. was the first person to make measurements on the angle between the ray in air and the normal to the plane refracting surface and the corresponding angle in water. In this book these angles will be called the **angle of inclination** in air and water respectively. He drew up tables from his measurements and enunciated the law that the angles of inclination in air and water are proportional to one another. This law is supported by his measurements for small angles, but it is quite wrong for large angles. There was very little serious scientific work for the next thousand years or so, but Alhazen first drew attention to the discrepancy between Ptolemy's results and his law. Snell was the first person to arrive at the correct statement of the law of refraction, although he never published his results. It is possible that Snell was led to his statement of the law from the known fact that a pond appears shallower than it really is, and so he came to measure the ratio $\frac{BD}{CD}$, assuming that the image of the point at the bottom of the pond was at C on the normal to the bottom (Fig. 1). It is possible that the constancy of this ratio for small values of the angle i led him to consider the ratio $\frac{AB}{AC}$, which is concerned with the incident and refracted rays themselves, and he found that this ratio is also constant for all angles of incidence. Descartes saw Snell's work, but he published the law of refraction in a more mathematical form. We see that Snell's original law states that $\frac{AB}{AC}$ is constant.

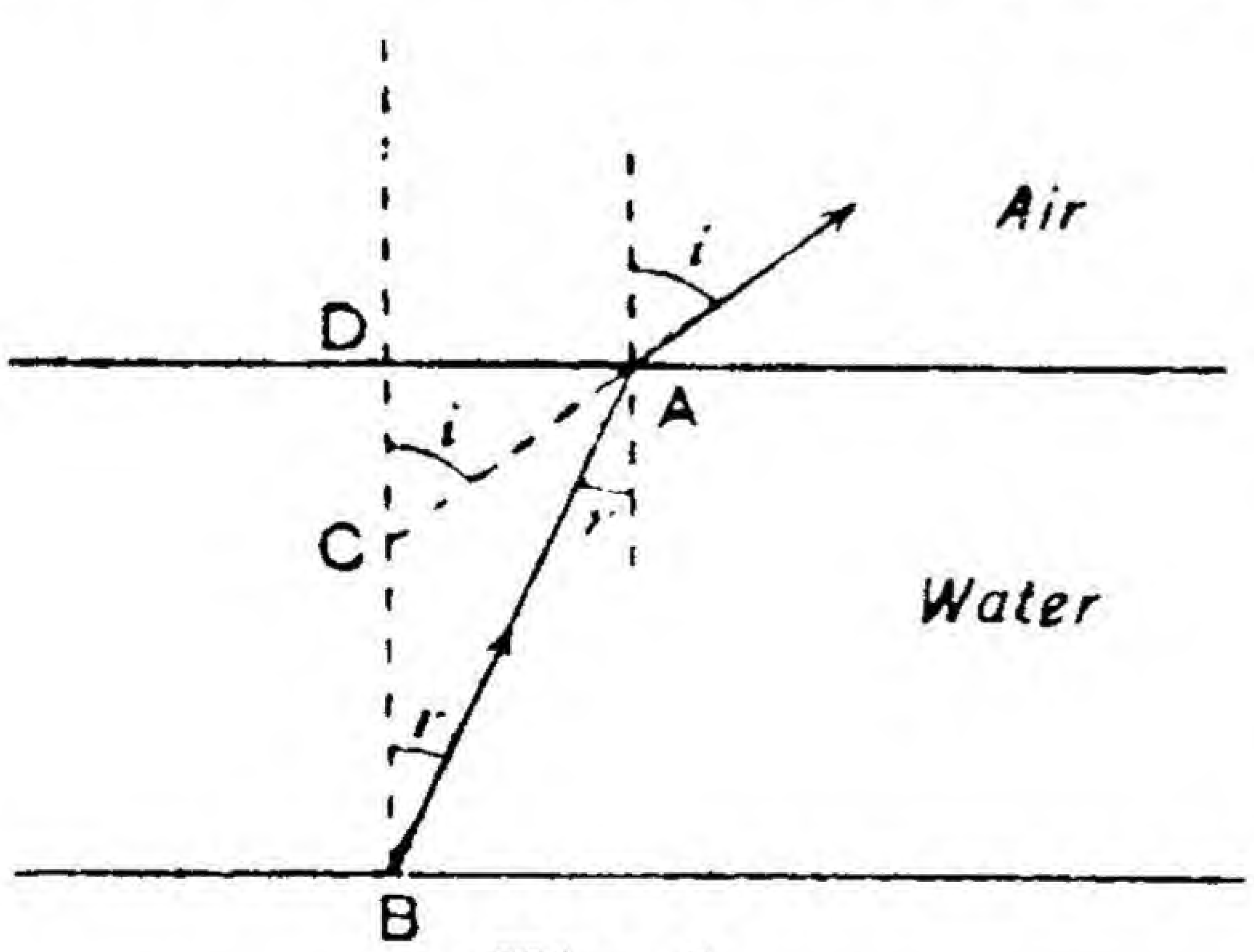


Fig. 1.

$$\therefore \quad \frac{\dfrac{AB}{AD}}{\dfrac{AC}{AD}} = \text{constant.}$$

$$\therefore \quad \frac{\operatorname{cosec} r}{\operatorname{cosec} i} = \text{constant, where } i \text{ and } r \text{ are the angles of inclination in air and water respectively,}$$

or $\dfrac{\sin i}{\sin r} = n$, where n is a constant for two given media.

This is Descartes' form of the law of refraction and n **is called the refractive index of water relative to air**, or, more usually, **the refractive index of water.** All refractive indices quoted in this book are to be taken as relative to air unless otherwise stated. So we have the modern statement of the law of refraction, which is commonly known as Snell's law, that **the refracted ray lies in the same plane as the incident ray and the normal to the refracting surface and the ratio of the sine of the angle of inclination in air to that in the medium is constant.** It is interesting that, although Descartes' form of the law is invariably used nowadays, Snell's original statement is the more convenient if we wish to construct the path of refracted rays so as to see quickly how the direction of the refracted rays will alter as that of the incident ray is changed. Let us suppose that we wish to study the refraction of light as it passes from air into glass. Take any point O (Fig. 2) on the refracting surface and draw with it as centre two circles whose radii are in the ratio 1·5, the refractive index of glass. To find the refracted ray corresponding to any incident ray AO, produce AO to meet the smaller circle at C and draw through C a line DCB normal to the refracting surface meeting the larger circle in B. Then, by Snell's own statement of his law, OB is the required refracted ray. We can see at once from this construction how there is a definite limit to the angle of refraction when the ray passes from air to glass and how there is a refracted ray corresponding to every possible incident ray. Now it is an experimental fact that, when a ray of light passing between two points is reversed, it retraces its path. Therefore it is equally clear that, if the light passes from glass to air, when the angle of inclination in glass exceeds a certain value there can be no corresponding ray in air. This occurs when the normal through B to the refracting

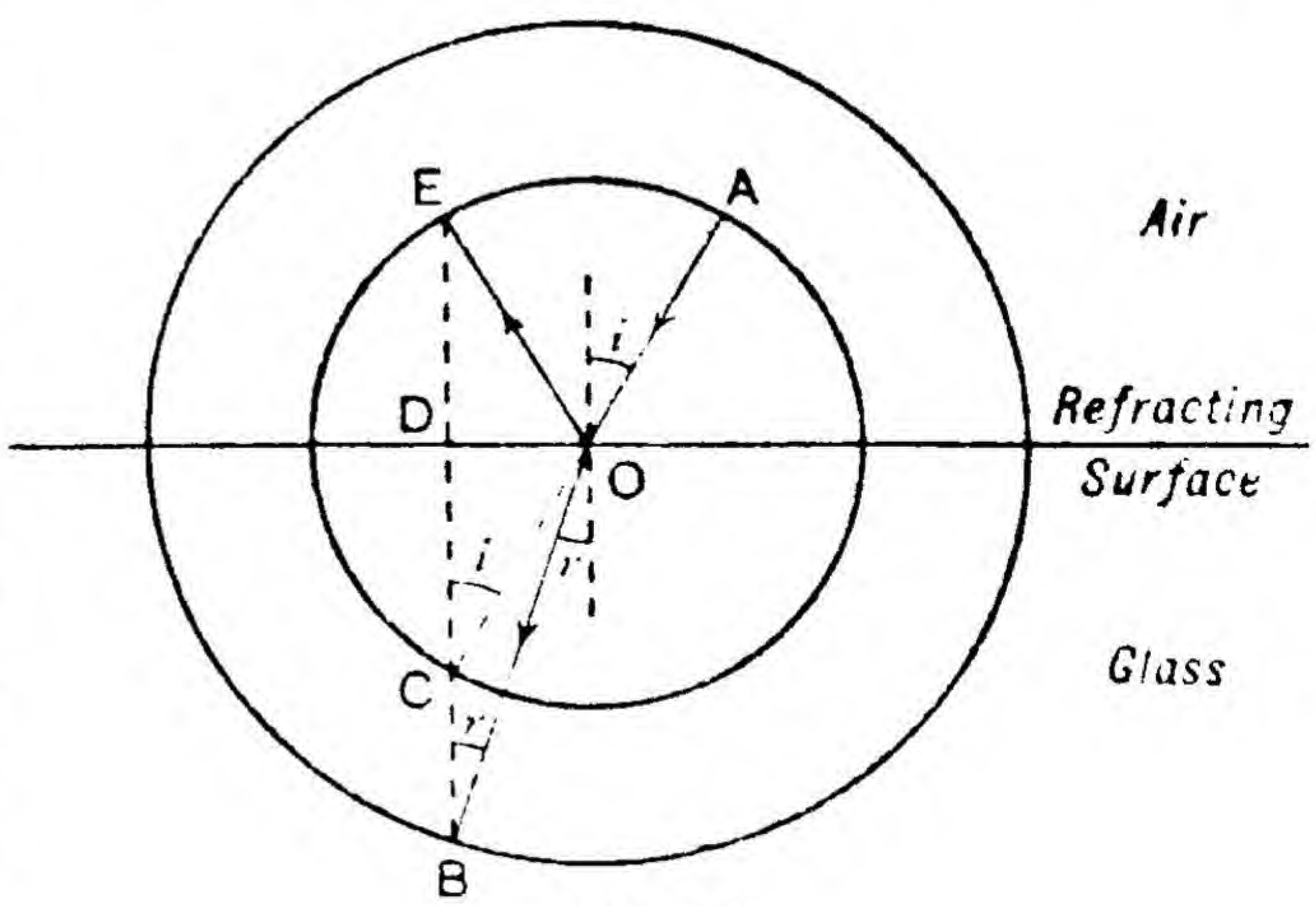


Fig. 2.

surface fails to cut the smaller circle. This merely indicates that Snell's law breaks down for such angles of refraction, and it remains to find out by experiment what occurs under these conditions. It is well known that all the light is reflected, the phenomenon being called **total reflection.** This breakdown of Snell's law can be equally well seen from its more familiar form. For $\sin i = n \sin r$ and when r exceeds the value r_c such that $n \sin r_c = 1$, then there is no possible value of i satisfying the equation. This merely indicates that Snell's law does not hold for values of $r > r_c$, which is known as the **critical angle**, but it does not show that no refraction is possible. This can only be decided by experiment. If the normal to the refracting surface is produced so as to cut the smaller circle at E again, then OE is the reflected ray. This construction suggests that *reflection is to be regarded as a particular case of refraction in which $n = -1$*, a point of view which can also be derived from the more usual form of Snell's law. This attempt to include reflection as a special case of refraction has been carried a stage further by Fermat, who succeeded in synthesising the three fundamental laws of light, rectilinear propagation, reflection, and refraction, into one general law, which we shall now consider.

4. FERMAT'S PRINCIPLE OF STATIONARY TIME

This law states that **the path actually taken by a ray of light in passing between two points is the path of least time**, that is, the time is a minimum compared to that which would be occupied in traversing other possible paths close to the actual path. We shall now show how the three fundamental laws of light can be derived from this principle. Rectilinear propagation follows at once; it is only true for an isotropic medium, in which the speed of light is the same in all directions. Consequently the path of least time between two points is the straight line joining them. The law of reflection follows from Fermat's principle in this way. Let us suppose that a ray of light is to go from A to B by way of the plane mirror PQ (Fig. 3) and let us consider any possible path ACB. If we draw AO perpendicular to the mirror and produce it to A′ so that AO=A′O, then AC=A′C whatever the position of C, and the length of the path ACB is A′C+CB. The path of least time between A and B is evidently the one of least length, since the light is travelling in the same medium throughout, and the path of minimum length is evidently ARB, where A′RB is a straight line, since A′RB is the line of least length between A′ and B. But the triangles ARO and A′RO are congruent, since AO=A′O, $\angle A\hat{O}R =$

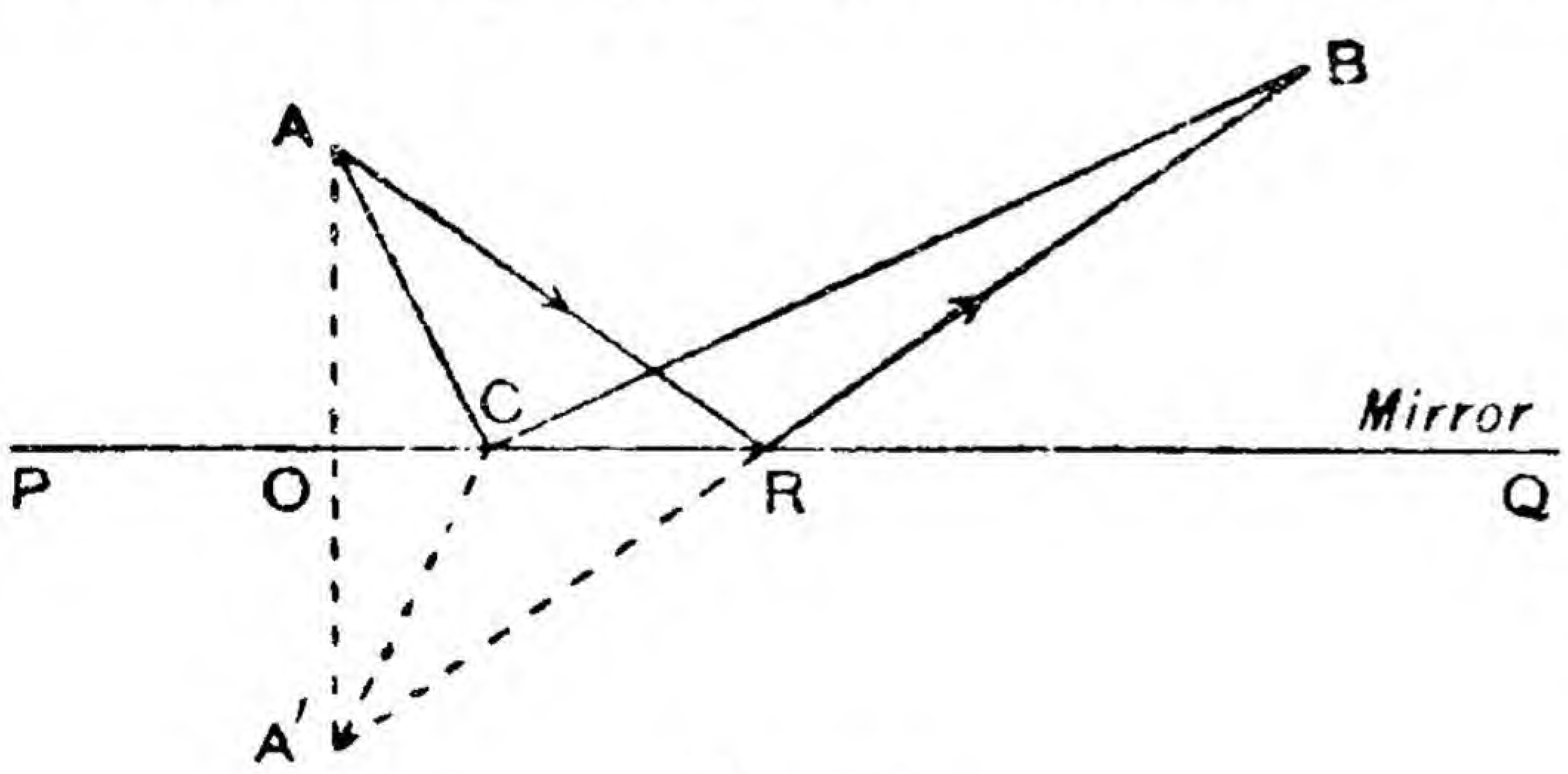


Fig. 3.

$\angle$A′OR, by construction, and OR is common.

$$\therefore \angle A\hat{R}O = \angle A'RO.$$

Also, since A′RB is a straight line.

$$\angle A'\hat{R}O = \angle B\hat{R}Q,$$

$$\therefore \angle A\hat{R}O = \angle B\hat{R}Q,$$

which is the law of reflection. The reader should notice carefully how the fact that A′RB is a straight line, which follows from Fermat's principle, is an essential part of this derivation of the law of reflection.

Finally, the law of refraction is also a consequence of Fermat's principle, as can be seen from a consideration of Fig. 4. Let ACB be the actual ray passing from a point A in air to a point B in glass, to consider two particular media for the sake of simplicity, and let ADB be a path very close to the actual path of the ray. It follows from Fermat's principle that the time occupied by the light in traversing ADB will, in the limit, be the same as that occupied in covering the actual path ACB. Draw CE perpendicularly to AC and DF perpendicular to BC. Then, as D approaches C, AE becomes more and more nearly equal to AC and BD equal to BF. It therefore follows from Fermat's principle that the time taken by the light to travel the distance DE in air is equal to that taken to travel the distance CF in glass. Then if c and c_g are the velocities of light in air and glass respectively, we have

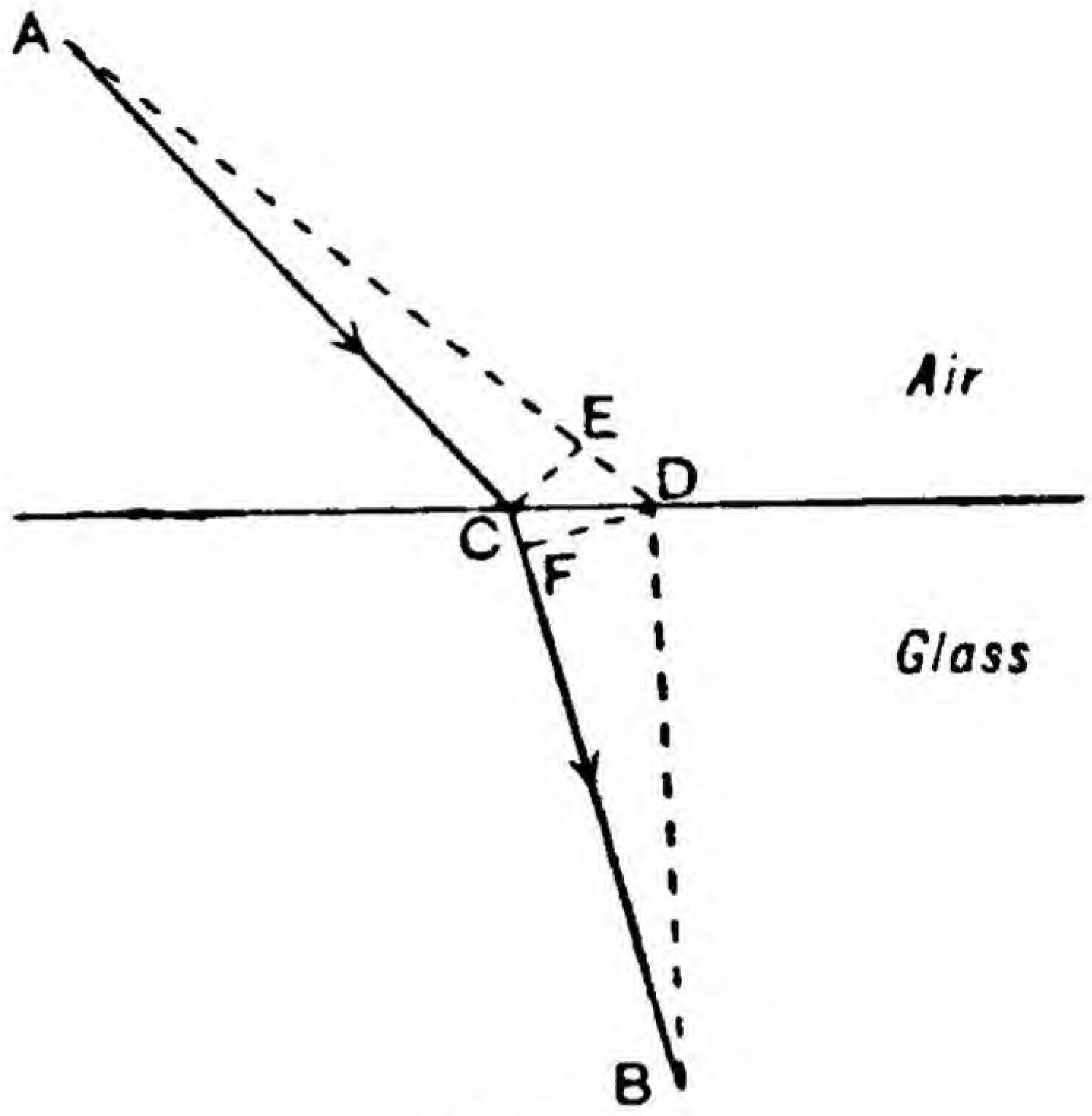


Fig. 4.

$$\frac{ED}{c} = \frac{CF}{c_g}$$

$$\therefore \frac{ED}{CF} = \frac{c}{c_g}$$

$$\therefore \frac{\sin E\hat{C}D}{\sin F\hat{D}C} = \frac{c}{c_g}, \text{ a constant.}$$

But $\angle E\hat{C}D = i$, the angle of inclination in air, and $\angle F\hat{D}C = r$, the corresponding angle in glass.

$$\therefore \frac{\sin i}{\sin r} = n, \text{ a constant,}$$

which is Snell's law. It should be noted that Fermat's principle not only leads to Snell's law, but that it gives a definite interpretation of the refractive index of a medium. It remains to be seen whether light is propagated with a finite velocity and, after that, if Fermat's prediction

that light travels more slowly in a dense medium such as glass than in air is true. Nevertheless his principle is a suggestive way of regarding the behaviour of rays of light. Its interpretation of the three fundamental laws of light makes a definite intellectual appeal; indeed, its appeal was so strong to thinkers of the Middle Ages that they regarded it as a stronger reason for believing the laws of reflection and refraction than the experimental results! We shall have occasion to refer to the principle later on, but it is worth while before leaving it to glance at its application to curved surfaces. Let us consider the possible paths taken by rays between the points A and B on a diameter ECD of a concave hemispherical mirror (Fig. 5), subject to the condition that the ray suffers only one reflection at the mirror. A and B are equidistant from C, the centre of curvature of the mirror, and CO $\perp$ ECD. It is clear from the law of reflection that there are three such paths, ADB, AEB, and AOB. Let us see if we can derive this result from Fermat's principle. Let APB be any possible path satisfying the above condition and let c be the radius of curvature of the mirror and AP$=a$, BP$=b$, and AC$=$BC$=d$. Since C is the middle point of the triangle APB, then, by Apollonius' theorem:

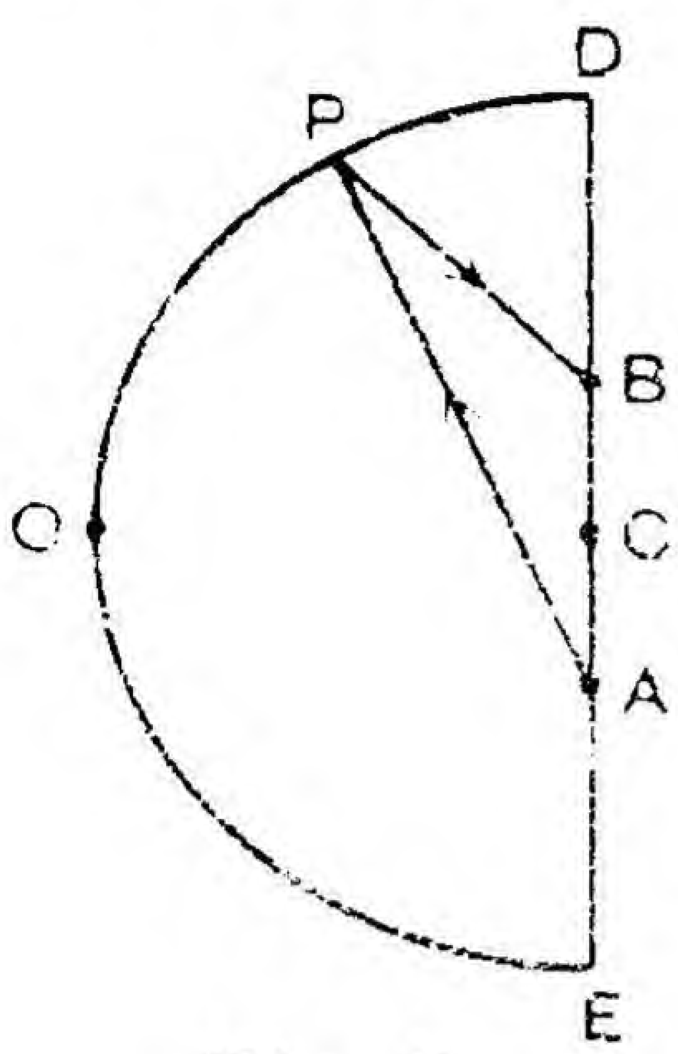


Fig. 5.

$$a^2+b^2=2(c^2+d^2)$$

$\therefore\ a^2+b^2=$constant for any position of P on the arc DOE.

Now $$(a+b)^2=2(a^2+b^2)=(a-b)^2$$

$$\therefore\ (a+b)^2=\text{constant}-(a-b)^2.$$

Therefore $(a+b)$ is a minimum when $(a\sim b)$ is a maximum. Now in the triangle APB $a<b+2d$ or $a-b<2d$. When P is at D or E, $a\sim b=2d$, its maximum value, and $(a+b)$ assumes its minimum value, which accounts for two of the paths. How are we to explain the remaining path AOB? When P is at O, $a=b$, $a-b=0$, its least value, and so $(a+b)$ is a maximum! In this case, then, the ray of light takes the path of greatest time! We must accordingly extend our original statement of Fermat's principle to include cases of this sort, which also occur in refraction at curved surfaces. If the path has to be either that of least or greatest time, then it is clear that the variation in the time taken among paths near to the actual path is in the limit zero, the usual condition for a maximum or minimum. In such cases we say that the time is **stationary**. So we call the principle "Fermat's principle of stationary time" and it states in its most general form that **a ray of light in passing between two points takes the path of least or greatest time.**

5. GEOMETRICAL OPTICS

Having established the two fundamental laws of rays of light we now turn to their consequences. This part of the subject is called **Geometrical**

Optics, which is the study of the reflection and refraction of rays of light at all types of surface and the application of the results to optical instruments. It does not concern itself with the nature of light, but only with those properties of rays which can be deduced from the two laws of reflection and refraction at plane surfaces. We can see the reason for this name, for these two laws are to be regarded as the "axioms" of this branch of geometry, and the various theorems relating to curved mirrors, refracting surfaces, prisms, and lenses are deduced from them just as the theorems of Euclid are deduced from a few axioms such as the shortest distance between two points is a straight line, parallel straight lines meet at infinity, and so on. We shall see that the results are at times somewhat abstract, as are those of Euclidean geometry, and that actual lenses and mirrors behave in a rather less simple way than our abstract lenses, in just the same way that the Euclidean straight line of length without breadth and depth cannot be realised in practice. But this does not affect the value of this kind of analysis of the properties of actual lenses and mirrors, and we shall see how it leads to a real understanding of the principles of such combinations of lenses and mirrors as telescopes and microscopes and therefore to their improvement.

We shall accordingly consider refraction at curved surfaces and the properties of lenses and then reflection at such surfaces and the properties of curved mirrors. Practical experience soon showed that the attempt to produce too much magnification by a single lens leads to coloured and blurred images, and it was this very trouble with refracting telescopes which led Newton to seek an explanation in the nature of white light itself. So we shall next consider his experiments on the splitting up of white light into the colours of the rainbow by a prism and the application of the results to the improvement of the images formed by single lenses. We shall see that results of a simple nature can only be obtained if we restrict ourselves to lenses, the diameter of whose circular outline is small compared to their focal length, and we shall next consider the type of effects obtained with lenses of large diameter, which are essential in certain optical instruments. It is natural to turn now to our own lens, the eye, and to consider its working and its defects. We shall deal with some problems of colour vision at this point also. The subject of vision naturally leads to the various aids to vision, such as the telescope, microscope, projection lantern, and other optical instruments; it also suggests the measurement of the luminous intensity of sources of light and the illumination of surfaces. Interest in these topics has been stimulated in recent years by the necessity for providing better artificial lighting in offices, factories, and on the roads. These improved methods in photometry, as this branch of the subject is called, have also been applied to the measurement of the intensity of spectral lines of various kinds and so have also found an application in pure science.

6. PHYSICAL OPTICS

So far we have established the properties of rays of light and their various applications, in short, we have developed the geometry of light rays. To put it in another way, we have obtained a classified set of experimental facts concerning light rays. On this basis we now proceed to the next stage in our quest, which is an attempt to answer the question: What is the nature of light? The theories of light and the experimental evidence bearing on them are usually known as **Physical Optics**, and the remainder of the book is devoted to this. There have always been two views as to the nature of light: the corpuscular theory, according to which a luminous body emits a stream of material particles which, on striking the retina, produce the sensation of sight; and the wave theory, which asserts that light is waves in some medium filling the whole of space. We shall review these two theories first of all in the light of those facts represented by Geometrical Optics which we have already established, and we shall see that neither can be regarded as entirely satisfactory, but that each suggests new lines of attack on the problem. The wave theory asserts that light will travel more slowly in water than in air, while the corpuscular theory asserts the opposite, and so this suggests a measurement of the speed of light, and indeed this was already of interest to settle the vexed question as to whether there was any propagation at all, for in what sense can the word be used if the speed of light is found to be infinite? So we next turn to the various measurements of the velocity of light and the results come down in favour of the wave theory. But this theory has still plenty of difficulties to overcome, as supporters of its rival are not slow to point out. They are waiting to see how the wave theory is going to explain away the rectilinear propagation of light, which so clearly distinguishes light from a well-known wave motion, sound. And when are the supporters of this theory going to demonstrate interference in light and what have they got to say about the "one-sidedness" of light as revealed in the experiments of Bartholinus on the double refraction of light by a crystal of calcite? So we turn to the phenomena of interference, diffraction, and polarisation, and we shall see that the supporters of the wave theory slowly but surely build up an impressive body of evidence in favour of their view, one by one removing the difficulties which stand in the way. We shall also see how this new knowledge, discovered almost exclusively for its own sake and for the intellectual satisfaction of fitting the facts into a rational scheme, has found a number of important industrial applications. So does new knowledge revolutionise the old trade, but the reader will also see throughout the book how science in its turn has had much help from and been inspired by the problems and needs of industry.

It seems as if the scheme is complete. One further problem remains: If light is waves, what kind of waves? What is the nature of the condition

which is propagated through the medium? Is it the deformation of some kind of elastic though non-material substance? Is it electrical in nature? The answers to these questions are beyond the scope of this book, but before we commence our journey of exploration in detail we must notice the clouds in the sky. As is usually the case, they are mere wisps to start with, giving no hint that they are soon to spread and cover the whole heavens. They have really been there for some time, but we were too busy with our successes to notice them. The photo-electric effect was discovered in 1888, and it was destined ultimately to defy explanation on the wave theory and to demand for its elucidation a new form of corpuscular theory. But we must not tarry further with these peeps into the future. Let us commence our journey!

EXAMPLES ON CHAPTER I

1. Show that it is possible to get an image on a screen by using a pin-hole instead of a lens. What are the advantages and the disadvantages of doing this? What are the conditions necessary to prevent the definition in the photograph being marred by the shape or size of the hole? (*Camb. Schol.*)

2. State the laws of reflection.
Apply these to find the length of the smallest mirror in which a man 5 ft. 10 in. high can see himself at full length. Why is a motor car driving-mirror convex? (*Oxford Schol.*)

3. Draw diagrams of the images obtained in two plane mirrors placed at an acute angle to one another, of an object placed between them not on the bisector of the angle.
Show how the laws of reflection at a plane surface may be used to explain the existence of a principal focus in a concave mirror of small aperture. (*Oxford Schol.*)

4. An even number of vertical mirrors are rigidly fastened together so that the angles between them do not vary. Show that the angular deviation of a horizontal beam of light reflected in succession from the mirrors is not affected by moving the mirrors.
Two vertical mirrors are inclined at an angle of 15° and a horizontal beam is incident at an angle of 60° on one of them and is reflected backwards and forwards between them; trace the path of the beam. (*Camb. Schol.*)

5. Two vertical mirrors are at right angles to each other. Show that a horizontal beam of light will be parallel to its original direction after it has suffered reflection from the two mirrors.
Describe the properties which such a system of mirrors would have. Can you suggest any applications for them? (*Camb. Schol.*)

6. Explain what is meant by the refraction of light, and describe one method by which the refractive index of a glass block could be measured.
A cube of glass of refractive index 1·5 stands on an ink mark on a sheet of paper. Why is it impossible to see the mark by looking through the vertical sides of the cube? (*Oxford Schol.*)

7. Calculate the angle of minimum deviation of light through a prism of refracting angle 60° and of material whose refractive index is 1·5. Also calculate the deviation when light is incident on one face of the prism at glancing incidence. (*Oxford Schol.*)

8. A ray of light passes nearly normally through a prism of *small* angle α, and refractive index μ. Show that the deviation δ is given by $\delta=(\mu-1)\alpha$.

A parallel beam of light falls normally on the first face of a prism of small angle, and undergoes a deviation of 1° 15′ in passing through the prism. The part of the beam which is reflected internally at the second face and emerges again at the first face makes an angle of 6° 31′ with the incident beam. Calculate the angle of the prism and the refractive index of the glass. (*Camb. Schol.*)

9. A small object is viewed normally through a parallel faced slab of transparent material of refractive index μ and thickness t. Find the apparent displacement of the object.

The base of a tank is a horizontal plate of glass 5 cm. thick of refractive index 1·6. On this is a layer of liquid of refractive index 1·5 and of thickness 10 cm., and on this liquid floats a layer of water 10 cm. thick of refractive index 4/3. An observer looking vertically downwards observes a spot on the lower side of the base. What is the apparent position of the spot ?

What would be the difference if the observer's eye were under the surface of the water ? (*Camb. Schol.*)

10. Light passes through the two equal faces of an isosceles prism and after emergence is incident on a mirror fixed to the base of the prism but extending beyond it. Show that the angle of minimum deviation is independent of the wave-length of the incident light and find its value. (*Oxford Schol.*)

11. Find an expression for the apparent movement of an object when a parallel-sided plate of refracting material is inserted between the object and the eye, the faces of the plate being perpendicular to the line of vision.

Decribe some practical method based on this phenomenon of determining the refractive index of a solid. (*O. and C.*)

12. State the conditions under which total reflection takes place. Describe how it can be applied in the construction of binoculars, and show that the refractive index of the glass must be greater than 1·41. (*Oxford. Schol.*)

13. An air cell consisting of two vertical parallel plates of glass separated by an air film can be rotated about a vertical axis in a liquid. Explain how you would use the apparatus to measure the refractive index of the liquid, and prove any formulæ you would use. (*Camb. Schol.*)

14. Two parallel-sided plates of glass are cemented together so as to enclose a thin film of air. Show that if the system is immersed in water, a ray of light will cease to be transmitted when its angle of incidence on the first glass surface is greater than the critical angle between water and air.

How can the principle of total reflection be applied to determine the refractive index of a solid ? (*O. and C.*)

15. If the refractive indices of water and glass are 4/3 and 1·5 respectively, find the critical angle for a glass water interface.

Describe a method, depending on the determination of the critical angle, of measuring the refractive index of a liquid. (*O. and C.*)

16. Give an explanation of the phenomenon of total reflection.

A ray of light travelling in a liquid is incident at the surface at the critical angle. A piece of plane parallel glass is placed in contact with the surface of the liquid and parallel to it. Show that the ray is incident on the glass air interface at the critical angle.

Describe a method of determining the refractive index of a liquid by the application of total reflection. (*O. and C.*)

17. What is meant by critical angle in relation to refractive media ? Describe carefully any experiment for the determination of the refractive index of a medium from observations on critical angles. A parallel beam of light from a cadmium arc is incident on a plane surface separating flint glass from air, passing first through the flint glass. Describe carefully what happens as the angle of incidence increases from 30° to 40°.

Refractive index of flint glass for cadmium red light = 1·645
Refractive index of flint glass for cadmium blue light = 1·665

(*London.*)

18. Give a statement of Fermat's Principle of Least Time and deduce the fundamental laws of reflection and refraction. (*London B.Sc.*)

19. Prove that when a beam of light is reflected from a rotating plane mirror, the angle turned through by the beam is twice that turned through by the mirror.

Explain fully how this fact is utilised (a) in the sextant, (b) in measuring, by an optical method, the refracting angle of a prism. Give diagram to illustrate each case. (*N.U.J.B.*)

20. State and explain Fermat's Principle of Least Time and use it to deduce the laws of reflection and refraction. Under what circumstances should the principle be replaced by one of maximum time ? (*Tripos, Part I.*)

Chapter II

REFRACTION AT CURVED SURFACES

7. INTRODUCTORY

Having established the two laws or axioms obeyed by rays of light, we now proceed to a logical deduction of their consequences. To put it in another way, we shall build up the theorems of geometrical optics. We shall deal with refraction at curved surfaces in the first place, partly because, as we have already seen, reflection is a particular case of refraction when $n = -1$, and so we may hope to get the theorems concerning reflection by making this simple substitution in the formulæ representing those of refraction; also refraction at curved surfaces may lead on to an understanding of lenses, which are more important than mirrors. We shall discuss spherical surfaces almost exclusively. It is natural to begin with these, because the sphere is the simplest surface there is, and so the results are likely to be easy to obtain and simple to interpret. But there is another reason for this choice. While we hope to find that our theorems are of interest in themselves, we shall also expect to be able to interpret with their aid the known behaviour of lenses, mirrors, and their combinations, such as telescopes and microscopes. And incidentally it is of interest that curved mirrors in the shape of polished metal surfaces and lenses in the form of a bottle of water were known and their simple properties had been elucidated before any theoretical interpretation was attempted. Now mirrors and the surfaces of lenses are nearly always spherical, since this is the only surface which can be accurately manufactured, so it is natural to begin by studying the effect of such surfaces on rays of light. This is one of the few cases in scientific investigation where the case of greatest practical importance is the one which lends itself to the simplest mathematical interpretation. We shall first extend Snell's law to refraction between any two media in order to give our results a more general validity.

8. REFRACTION BETWEEN ANY TWO MEDIA

If a ray of light AB is incident on a plane surface separating two media of refractive index n_1 and n_2 (Fig. 6), the ray in the second medium BC will bend towards the normal if n_2 is greater than n_1. If measurements are made of the angles of inclination i_1 and i_2 in the two media over a

suitable range, it is found that the results satisfy the following law :

$$\frac{\sin i_1}{\sin i_2}=\frac{n_2}{n_1}$$

which may be put in the more symmetrical form

$$n_1 \sin i_1=n_2 \sin i_2 \quad . \quad . \quad (1)$$

We shall always use this form of the law of refraction in what follows and it is to be emphasised that it is based on experimental fact and that it is a more general form of the law of refraction than Snell's law. Indeed, it reduces to that law, if the first medium is assumed to be air, when $n_1=1$ and equation (1) reduces to the ordinary statement of the Snell's law.

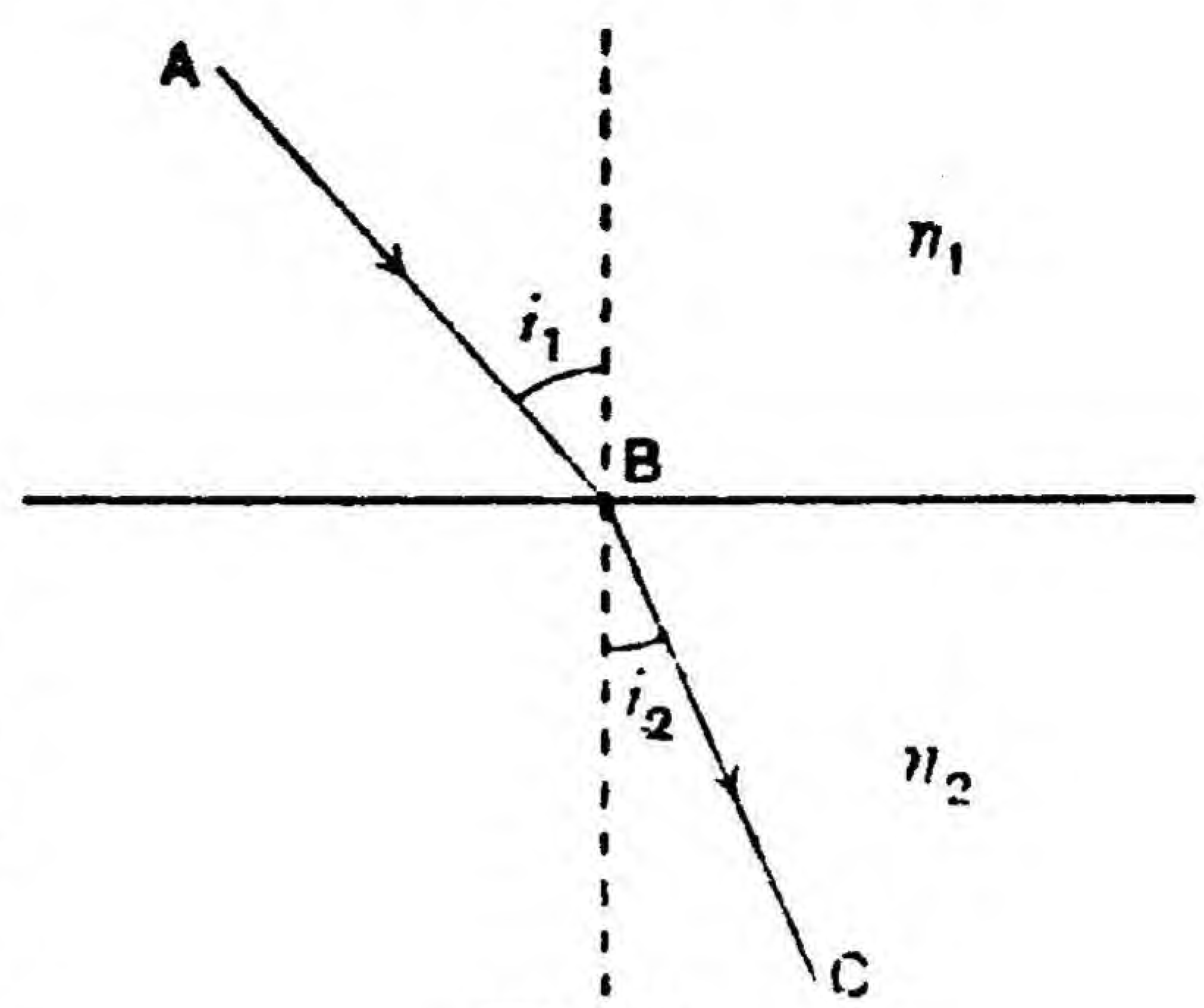


Fig. 6.

9. REFRACTION AT A SINGLE SPHERICAL SURFACE

Case I. A convex surface producing a real image of a real object.

Let A be a point object in a medium of refractive index n_1 separated from one of refractive index n_2 $(n_2>n_1)$ by a spherical surface of centre C radius c (Fig. 7). Let the line AC cut the surface at O, called the pole of the surface, the line CO being its axis. We will investigate whether

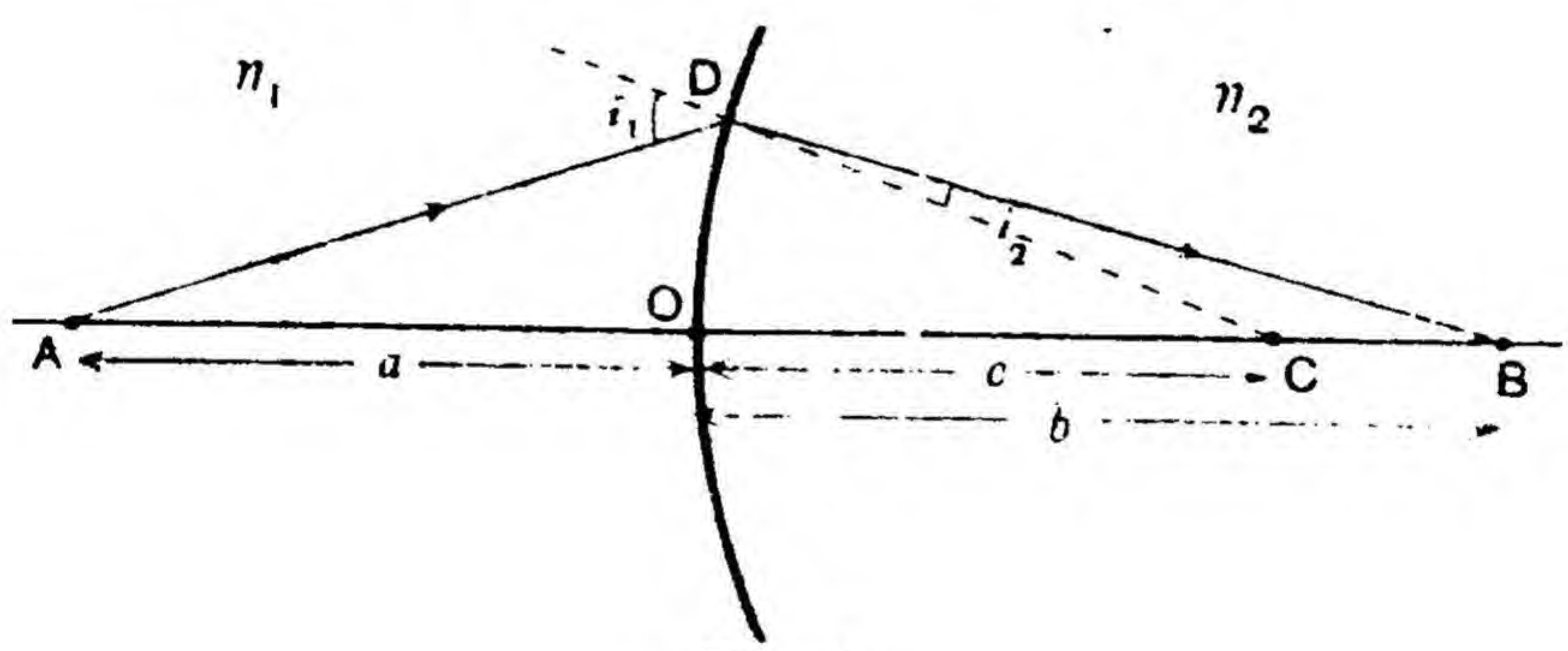


Fig. 7.

the spherical surface forms a point image of A and, if so, where it will be. Draw any ray AD from A cutting the surface at D and let the ray in the second medium cut the axis at B. We have to see if the position of B is independent of the angle between AD and the axis of the surface. Now the area of $\triangle$ ADB=area of $\triangle$ ADC+area of $\triangle$ CDB.

If AD$=p$ and BD$=q$

$$\tfrac{1}{2}\, pq \sin (i_1-i_2)=\tfrac{1}{2}\, pc \sin i_1+\tfrac{1}{2}\, qc \sin i_2$$

$$\therefore\ pq (\sin i_1 \cos i_2-\cos i_1 \sin i_2)=pc \sin i_1+qc \sin i_2$$

By the law of refraction,

$$n_1 \sin i_1=n_2 \sin i_2$$

Eliminating i_1 from these two equations, we have

$$pq\left(\frac{n_2}{n_1}\sin i_2 \cos i_2 - \cos i_1 \sin i_2\right) = pc\,\frac{n_2}{n_1}\sin i_2 + qc \sin i_2$$

$$\therefore\ pq(n_2 \cos i_2 - n_1 \cos i_1) = pc\, n_2 + qc\, n_1$$

Dividing out by pqc, we have

$$\frac{n_2}{q} + \frac{n_1}{p} = \frac{n_2 \cos i_2 - n_1 \cos i_1}{c} \quad . \quad . \quad . \quad (2)$$

It is not easy to see at a glance from this equation if the position of B is independent of the angle which AD makes with the axis, but it does not appear likely. We will postpone a definite consideration of this point until later. But, *if* we restrict ourselves to rays which make only a small angle with the axis, we have $\cos i_1 = \cos i_2 = 1$ and $p=a$ and $q=b$ (Fig. 7), and equation (2) becomes

$$\frac{n_2}{b} + \frac{n_1}{a} = \frac{n_2 - n_1}{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

It should be emphasised that these angles are not *bound* to be small; indeed, in the case of high-power microscopic objectives, for example, the rays from the object striking the edge of the objective make large angles with the axis. We have merely made this restriction in order to get a simple mathematical expression, which we can easily interpret. It remains to be seen if these restrictions are ever realised in practice. Rays satisfying this restriction are called **paraxial rays.** Again it should be stressed that the statement $\cos i_1 = 1$ is not mathematically true; it merely asserts that i_1 is so small that $\cos i_1$ differs from 1 by a quantity less than the experimental error made or allowed in our experiments.

It follows from equation (3) that, for given values of a, c, n_2, and n_1, there is a definite value of b independent of i_1 and therefore of the angle which the ray AD makes with the axis of the surface provided it is small. That is, all paraxial rays from A in the plane of the diagram pass through B after refraction. Since the surface is symmetrical about its axis, it follows that the paraxial rays from A in any other plane containing the axis will pass through B after refraction. Therefore all paraxial rays from A pass through B after refraction; in other words, B is the image of A. We must now see if this is true for other positions of the object and for concave surfaces.

Case II. A convex surface producing a virtual image.

Our problem is just the same in this case, in which we use the same notation as before. We have from Fig. 8,

(area of $\triangle$ ADB = area of $\triangle$ BDC − area of $\triangle$ ADC)

$$\therefore\ \tfrac{1}{2}pq \sin(i_1 - i_2) = \tfrac{1}{2}qc \sin i_2 - \tfrac{1}{2}pc \sin i_1$$

Also, by the law of refraction,

$$n_1 \sin i_1 = n_2 \sin i_2$$

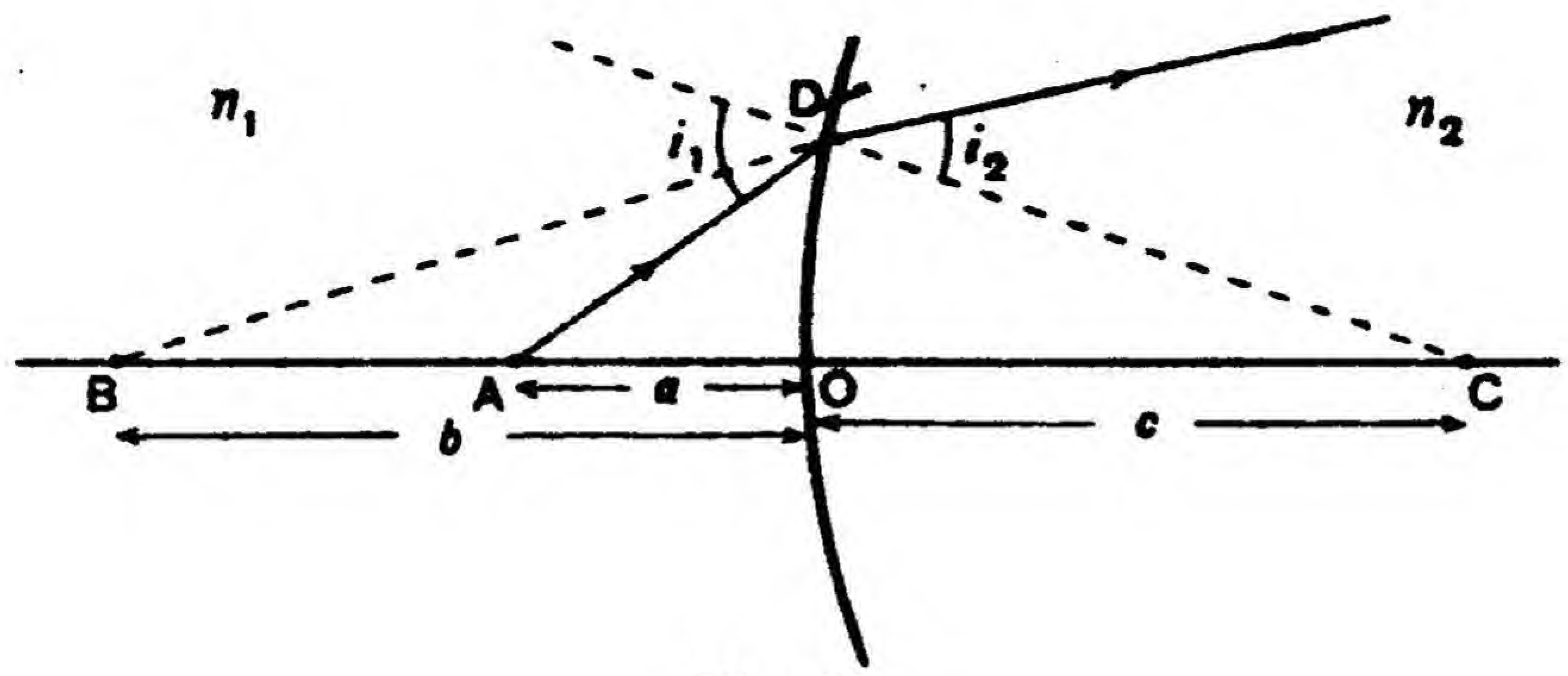


Fig. 8.

From these two equations we have, just as before,

$$pq\left(\frac{n_2}{n_1}\cos i_2 - \cos i_1\right) = qc - pc\frac{n_2}{n_1}$$

$$\therefore \quad pq(n_2 \cos i_2 - n_1 \cos i_1) = qc\, n_1 - pc\, n_2$$

$$\therefore \quad \frac{n_1}{p} - \frac{n_2}{q} = \frac{n_2 \cos i_2 - n_1 \cos i_1}{c}$$

Therefore, for paraxial rays,

$$\frac{n_1}{a} - \frac{n_2}{b} = \frac{n_2 - n_1}{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

So again we see that the value of b for given values of a, c, n_2, and n_1 is the same for all paraxial rays in the plane of the diagram and by symmetry in every plane through the axis of the surface, and so B is the point image of the point object A on the axis of the system. It must be emphasised that this is only true for rays making small angles with the axis.

Case III. A concave refracting surface.

In the same way as in the previous two cases, we have from Fig. 9,

area of $\triangle$ ADB = area of $\triangle$ ADC − area of $\triangle$ BDC

$$\therefore \quad \tfrac{1}{2} pq \sin(i_1 - i_2) = \tfrac{1}{2} pc \sin i_1 - \tfrac{1}{2} qc \sin i_2.$$

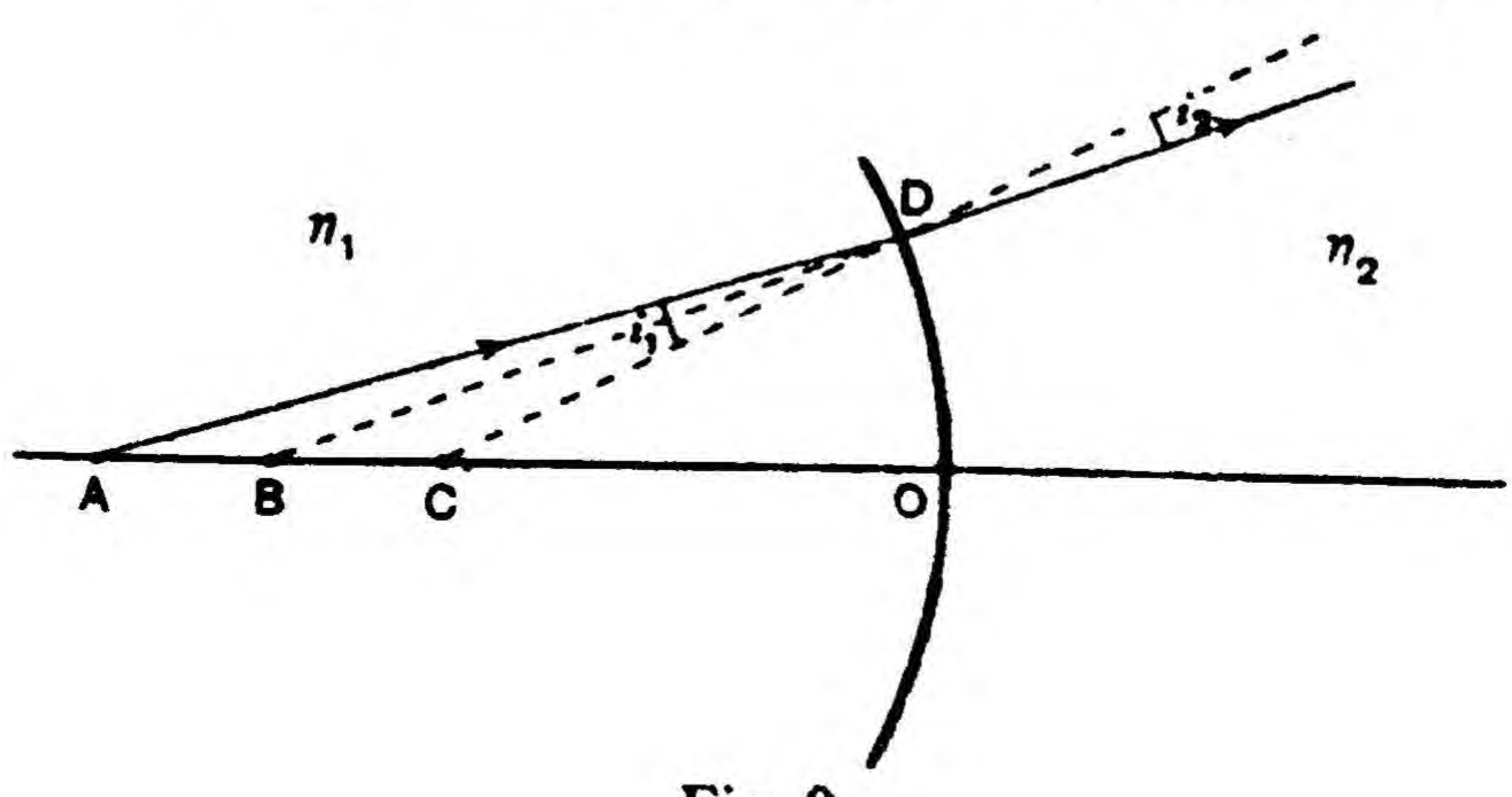


Fig. 9.

And from the law of refraction and this equation, we have

$$pq\,(n_2 \cos i_2 - n_1 \cos i_1) = pc\,n_2 - qc\,n_1$$

$$\therefore \quad \frac{n_2}{q} - \frac{n_1}{p} = \frac{n_2 \cos i_2 - n_1 \cos i_1}{c}$$

or, for paraxial rays,

$$\frac{n_2}{b} - \frac{n_1}{a} = \frac{n_2 - n_1}{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

As before, then, the concave refracting surface forms a virtual point image B on its axis of a point object A on its axis, if only paraxial rays are allowed to pass through the surface. It is already apparent that some definite restrictions will have to be placed on actual lenses, for example, if they are to conform with the conditions under which our results are valid. So we have our first important result, which we shall call theorem 1.

Theorem 1. **Any spherical refracting surface forms a point image on its axis of a point object on the axis if the surface is stopped down so that only paraxial rays can pass through it.**

10. THE CONVENTION OF SIGNS

If the results we have obtained as expressed in equations (3), (4), and (5) are examined, we see that their form is the same, but that they differ just in the matter of the *sign* prefixed to each term. The present position, then, is that we have three equations each applying to its own case and each differing from the other only in a matter of sign. We cannot be satisfied with such a clumsy set of results, the reason for which is not difficult to understand and will show us how to express them in a more elegant form. Our equations are expressed in distances, which are pure numbers without a sign. Therefore we cannot introduce into our equation on which side of the pole of the surface the image is situated but only its distance away and *so we cannot distinguish between a real and a virtual image* ; so we have to have one equation satisfying the formation of real images and one for virtual images. Again, we cannot introduce into our equation the difference between a concave and convex surface, and so we have to have another equation for concave surfaces. If we express our equation in such quantities that we can distinguish between real and virtual images, convex and concave surfaces, we should expect to get a single equation satisfying all the cases. To put it in another way, only if we introduce all the physical conditions into our equation can we expect to get a complete solution of the problem from it. One way of doing this is to express our equations in terms of the **position** or co-ordinate of the object, image, and centre of curvature of the surface. Since these points are all on the axis of the surface, we can specify their **position** by giving their *distance* from the pole of the surface and prefixing a *sign* to that distance to indicate on which side of the pole the point is situated. A number of different conventions to determine this

sign have been used in the past and some confusion has thereby been produced, but we shall use in this book one of two conventions recommended by a recent committee appointed by the Physical Society of London to consider the teaching of Geometrical Optics. Taking the pole of the refracting surface as the origin, distances measured in the *same* direction as the *initial* direction of the light will be counted *positive* and those measured in the *opposite* direction will be counted *negative.* On this convention, it will be seen that the *position* of the object in all the above cases is $-a$ units, since it is a units from the origin in the opposite direction to that in which the light is travelling. This will be the case whether the incident light goes from left to right or right to left. In this book we shall always draw the incident light from left to right so that this convention also agrees with the usual co-ordinate geometry convention for co-ordinates or positions. Again, the position of the centre of curvature of the convex surface considered above is $+c$, whereas that of the concave surface is $-c$. We now define the radius of curvature of a surface as the position of its centre of curvature relative to its pole and so its complete specification involves a sign as well as a magnitude. Finally the position of the real image, considered above, is $+b$, while that of the virtual image is $-b$, so we can distinguish between real and virtual images.

Let us now rearrange equations (3), (4), and (5) to be in terms of positions instead of distances. Throughout this book, to avoid confusion, the letters u, v, r, and f will be used to denote the **position** of the object, image, centre of curvature, and focus of a system, while a, b, c, and d will be used to denote their distance from the origin.

For Case I, we have $u=-a$, $v=+b$, $r=+c$.

$$\therefore \quad a=-u,\ b=v, \text{ and } c=r.$$

Substituting these values of a, b, and c in equation (3), we have

$$\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

For Case II, we have $u=-a$, $v=-b$, $r=+c$.

Substituting the values of a, b, and c from these equations in equation (4), we have again

$$\frac{n_1}{-u}-\frac{n_2}{-v}=\frac{n_2-n_1}{r}$$

or

$$\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{r}$$

For Case III, we have $u=-a$, $v=-b$, $r=-c$, and substituting these values of a, b, and c in equation (5), we have once more

$$\frac{n_2}{-v}-\frac{n_1}{-u}=\frac{n_2-n_1}{-r}$$

or

$$\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{r}$$

So, as we expect, by working with **positions** instead of distances, all the possible cases of refraction at one spherical surface are satisfied by one equation and the data given, when substituted in that equation, will give a complete answer to the unknown quantity.

The use of this so-called convention of signs sometimes presents difficulties to the student, so it will be further illustrated by an actual example. Suppose a beam of rays parallel to the axis falls on a convex lens made of glass of refractive index 1·5, the radius of curvature of each face of the lens being 20 cm. Find at what point on the axis of the lens the paraxial rays will come to a focus, assuming that the lens is of negligible thickness and is situated in air. For the first surface of the lens, we have $n_1=1$, $n_2=1\cdot5$, $r=+20$, $u=-\infty$. Therefore to find v, we have from equation (6)

$$\frac{1\cdot5}{v}-\frac{1}{-\infty}=\frac{1\cdot5-1}{+20}$$

$$\therefore \quad \frac{1\cdot5}{v}=\frac{0\cdot5}{20}$$

$$\therefore \quad v=+60 \text{ cm.}$$

Notice that when substituting in the general equation we do not give v a sign, as we do to u and r. The reason is that v, the unknown quantity, has both magnitude *and* sign and the equation will give both. Some readers make the mistake of inserting the sign for v as well as for u and r and then they wonder why the equation does not give the correct sign! It should be evident that this is wrong if v is the unknown; if it is unknown *both* its sign and magnitude are unknown and the equation will give both. So we see that the rays, after entering the first lens, are *converging* on a point 60 cm. from the centre of the lens. Such a point is called a **virtual object**, since the rays never reach it, suffering another refraction at the second surface of the lens before they can do so. For this refraction, $n_1=1\cdot5$, $n_2=1$, $u=+60$ cm. as the lens is of negligible thickness, $r=-20$ cm.

To find v, we have

$$\frac{1}{v}-\frac{1\cdot5}{+60}=\frac{1-1\cdot5}{-20}$$

$$\therefore \quad v=+20 \text{ cm.}$$

So the paraxial rays of the beam come to a focus 20 cm. from the centre of the lens on the opposite side from that at which the light started. In other words, the lens is a converging one of focal length 20 cm. It should be emphasised that we have, as yet, no knowledge of where rays making finite angles with the axis will come to a focus, if indeed they come to a point focus at all.

It may be as well to add one further point about the convention of signs. There is no necessity to use it; the alternative is to use a different

equation for each different case and to remember which of a number of very similar equations applies to which case. But the one equation relating to positions embraces all the physical conditions of every case and so satisfies them all, and it is more elegant and simpler to use this one equation and for that reason we shall, in future, express our final equations in terms of positions. Accordingly we shall re-write the equations for rays making any angle with the axis in positions, when the reader can verify that it becomes in every case

$$\frac{n_2}{q}-\frac{n_1}{p}=\frac{n_2 \cos i_2-n_1 \cos i_1}{r} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

if q and p are given signs in accordance with the above convention.

11. INTERPRETATION OF THE RESULTS

Let us now look at some of the more interesting consequences of equation (6). If we start with convex surfaces, we may put $r=+c$; putting $u=-\infty$, and the corresponding value of $v=f_2$,

$$\frac{n_2}{f_2}=\frac{n_2-n_1}{c}$$

or

$$f_2=\frac{c\, n_2}{n_2-n_1}$$

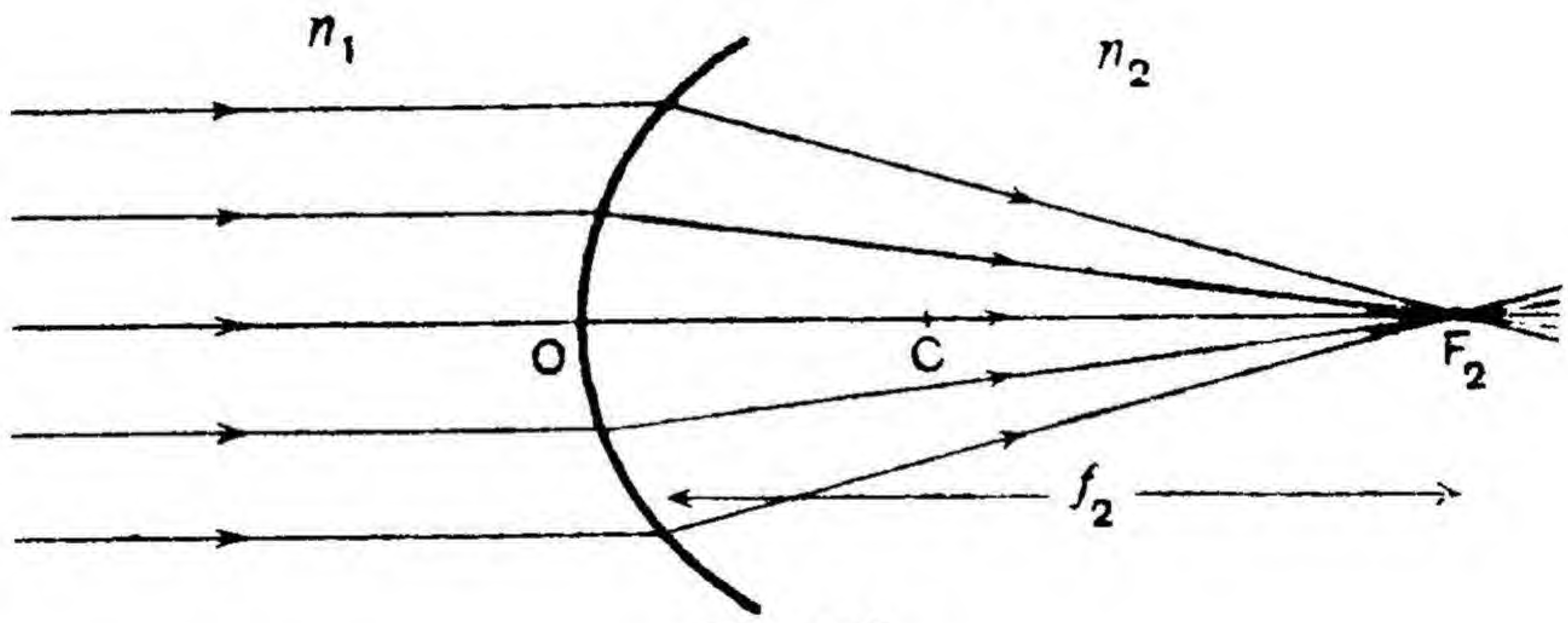


Fig. 10.

So we see that all rays parallel to the axis of the surface come to a focus at a point F_2 distant $\frac{c\, n_2}{n_2-n_1}$ from the pole of the surface (Fig. 10). This is called the **second principal focus** of the surface and the **second focal length, f_2, is its position relative to the pole of the surface.** Again, if the object is so placed that the rays from it are all parallel to the axis of the surface after refraction, the object is at the first principal focus of the surface F_1 and its position relative to the pole is called the first focal length, f_1.

$$\therefore \quad \frac{n_2}{+\infty}-\frac{n_1}{f_1}=\frac{n_2-n_1}{+c}$$

$$f_1=-\frac{n_1\, c}{n_2-n_1}$$

So the first principal focus of a convex surface is $\frac{n_1 c}{n_2 - n_1}$ units from the pole on the same side as that from which the incident light started (Fig. 11). Notice that $f_1 \neq f_2$.

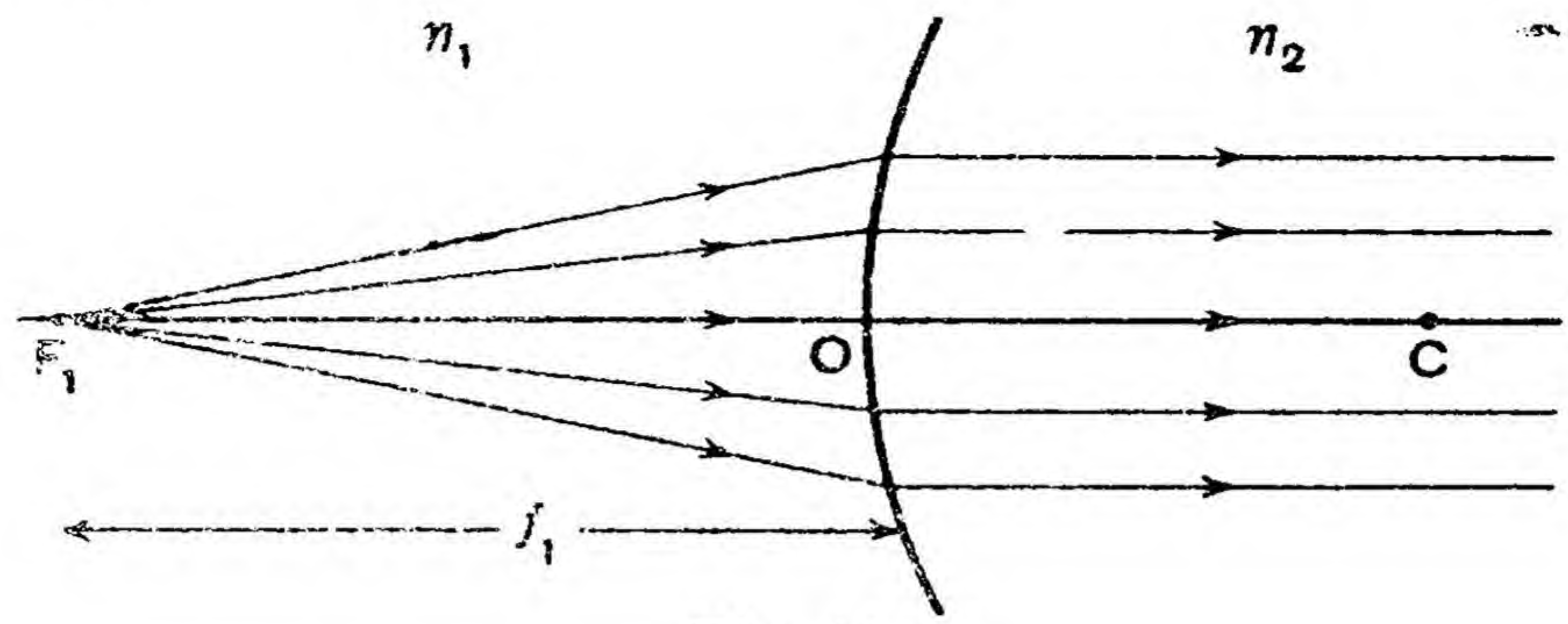


Fig. 11.

If $u = -a$,

$$\therefore \quad \frac{n_2}{v} - \frac{n_1}{-a} = \frac{n_2 - n_1}{+c}$$

$$\therefore \quad \frac{n_2}{v} = \frac{n_2 - n_1}{c} - \frac{n_1}{a}$$

There are two cases of interest. If a is greater than the numerical value of f_1, then $\frac{n_1}{a}$ is less than $\frac{n_2 - n_1}{c}$ and so v is positive; that is, if a point object is further from the surface than the first principal focus, the surface produces a real image of it. Conversely, if the object is inside the principal focus, then a is less than the numerical value of f_1 and so v is negative and the surface forms a virtual image of the object (Fig. 12).

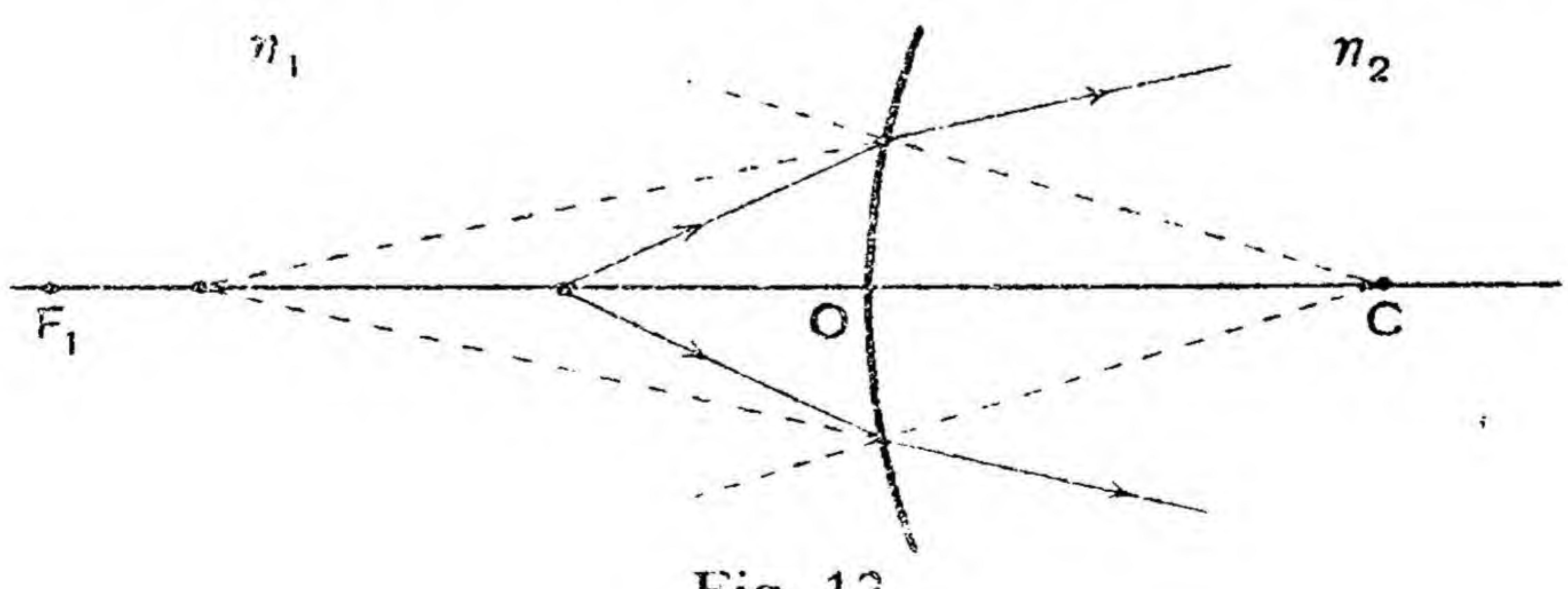


Fig. 12.

The case of the concave surface can be treated along similar lines. Here we have $r = -c$ and when $u = -\infty$,

$$\frac{n_2}{v} - \frac{n_1}{-\infty} = \frac{n_2 - n_1}{-c}$$

$$\therefore \quad v = -\frac{n_2 c}{n_2 - n_1}$$

In other words, if a beam of rays parallel to the axis strikes the surface,

they appear after refraction to diverge from a point F_2 on the axis a distance $\frac{n_2 c}{n_2-n_1}$ from the pole of the surface (Fig. 13). This point is the

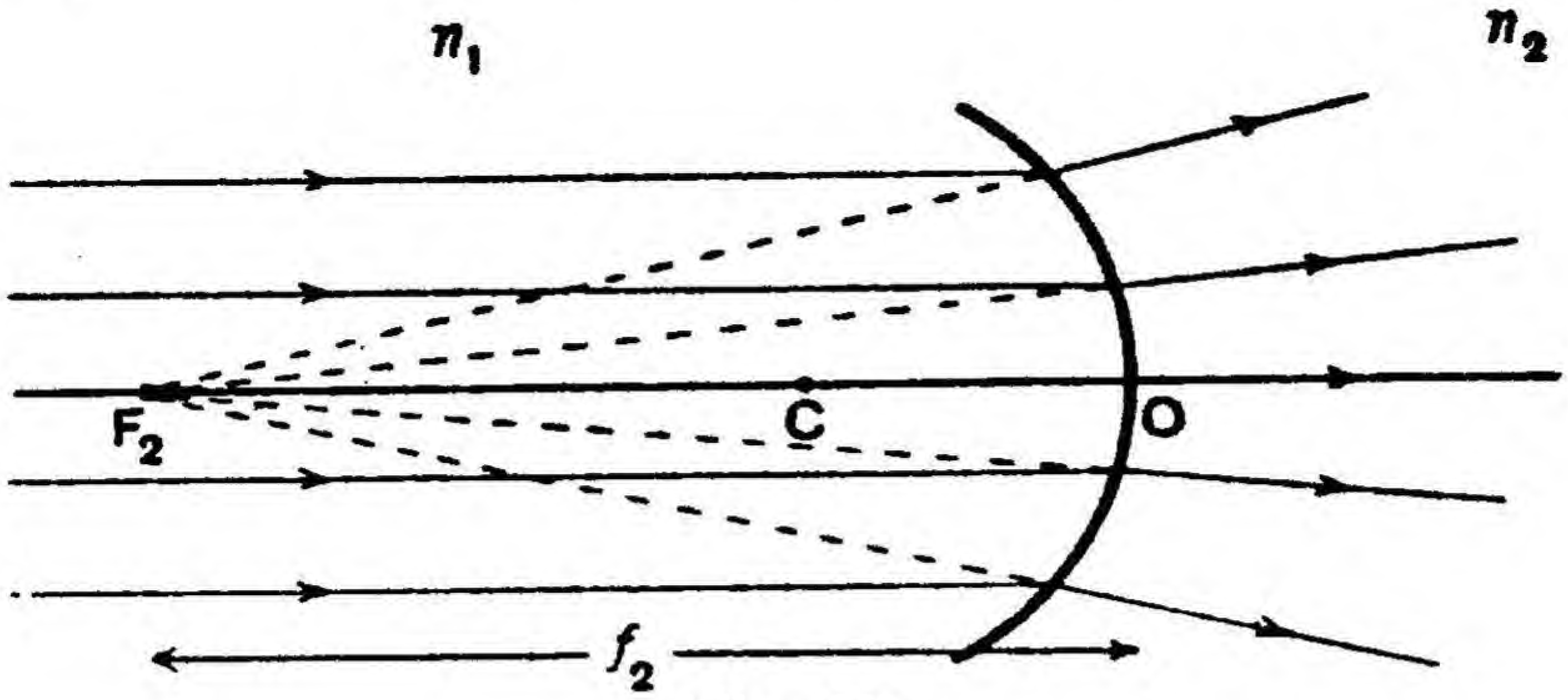


Fig. 13.

second principal focus of the surface, its position relative to the pole being the second focal length, f_2. In the same way, if $v=+\infty$,

$$\frac{n_2}{+\infty}-\frac{n_1}{u}=\frac{n_2-n_1}{-c}$$

$$\therefore \quad u=+\frac{n_1 c}{n_2-n_1}$$

and so if a beam of rays is converging to a point F_1 distant $\frac{n_1 c}{n_2-n_1}$ from the pole of the surface on the opposite side to that from which the light is coming, it is rendered parallel to the axis after refraction (Fig. 14).

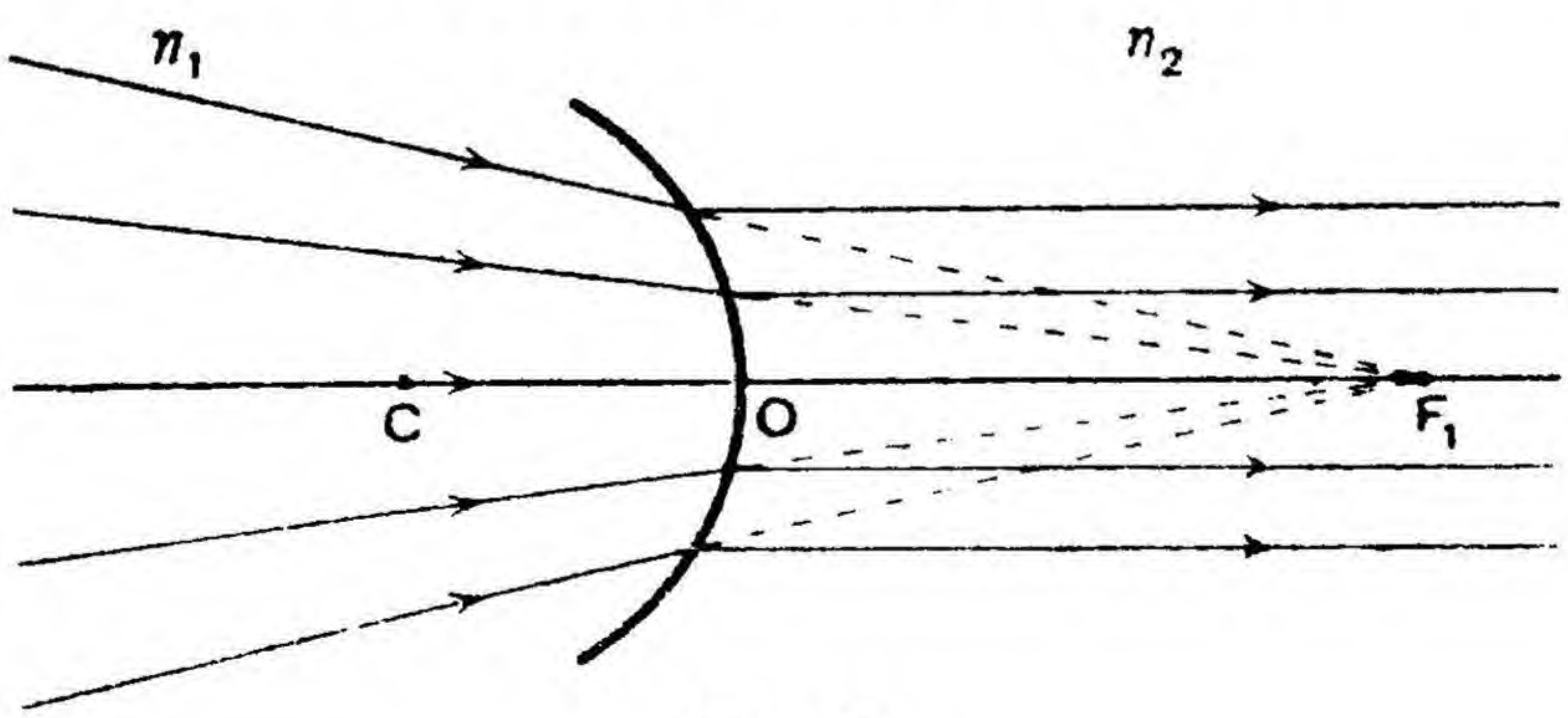


Fig. 14.

The point to which the rays converge before refraction is called the first principal focus of the surface and its position relative to the pole the first focal length, f_1. Lastly, if $u=-a$, where a may have any value,

$$\frac{n_2}{v_1}-\frac{n_1}{-a}=\frac{n_2-n_1}{-c}$$

$$\therefore \quad \frac{n_2}{v_2}=-\left(\frac{n_2-n_1}{c}+\frac{n_1}{a}\right)$$

So v must be negative, which means that a concave refracting surface

must form a virtual image of a real object. But if we have a virtual object at A, then $u=+a$ and in this case, if a is less than the numerical value of f_1, a real image is formed at B, as the reader should deduce for himself (Fig. 15). We may conclude this account of the properties of spherical refracting surfaces by a reference to two points of importance. In all the diagrams in which the formation of an image is illustrated, only two rays are drawn from the object and shown passing through or appearing to diverge from the image. This is merely done for the sake of simplicity and is justified because our analysis *proves* that *all* paraxial rays from a point object on the axis do pass through one point also on the axis after refraction. It cannot be too strongly emphasised that the finding of the point where two rays from the object meet after passing through an optical system does not locate the image of that object formed

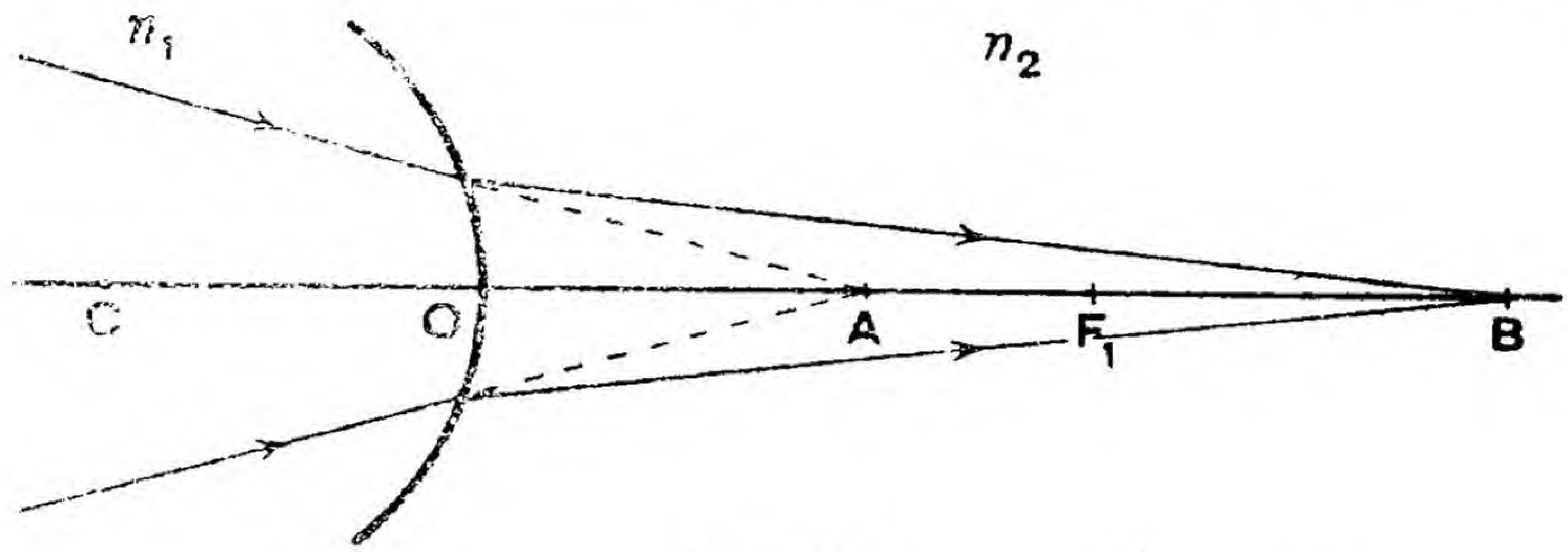


Fig. 15.

by that system, for it does not prove that a point image has been formed at all. To do this, it is necessary to show that all the rays from the object passing through the given system pass through the same point (or appear to diverge from the same point) on emerging from the system. Secondly, we cannot remind ourselves too often that the conclusions we have reached are only true for paraxial rays, and so we are really dealing with an ideally simple type of surface ; we have definitely restricted ourselves to this abstract type of surface to start with ; we will see how far actual systems approximate to our abstract system later and then we can decide if we have to investigate systems approximating more closely to those used in practice.

So far, then, we have proved from the law of refraction that a spherical refracting surface forms a point image of a point object on its axis if only paraxial rays are allowed to pass through the surface. Our second theorem follows at once.

Theorem 2. Any number of co-axial spherical refracting surfaces form a point image of a point object on the axis.

12. OBJECTS OF FINITE SIZE

We have seen that our analysis only applies to an abstract system, which means a system which cannot be realised in practice owing to

certain restrictions or qualifications which are specified in order that it shall have ideally simple properties. Other examples of abstract conceptions are the point in geometry, something which has no dimensions but only position; it is impossible to make a point or to perceive a point by the senses, but it is possible to think of a point or to imagine one. It is this impossibility of making but possibility of imagining or thinking of a concept which makes it abstract. Another example is the ideal gas in physical science; it is impossible to realise it in practice, but we can think of it and even specify its properties. Now let us turn to the nature of the object itself. We have so far only considered a point, which, as we have just seen, cannot be realised in practice. But we can analyse an object of finite size into a series of points and what will a spherical refracting surface do to such an object? Let A be a point object on the axis and B be its real image formed by a convex spherical surface (Fig. 16). Let the whole diagram be rotated through a *small*

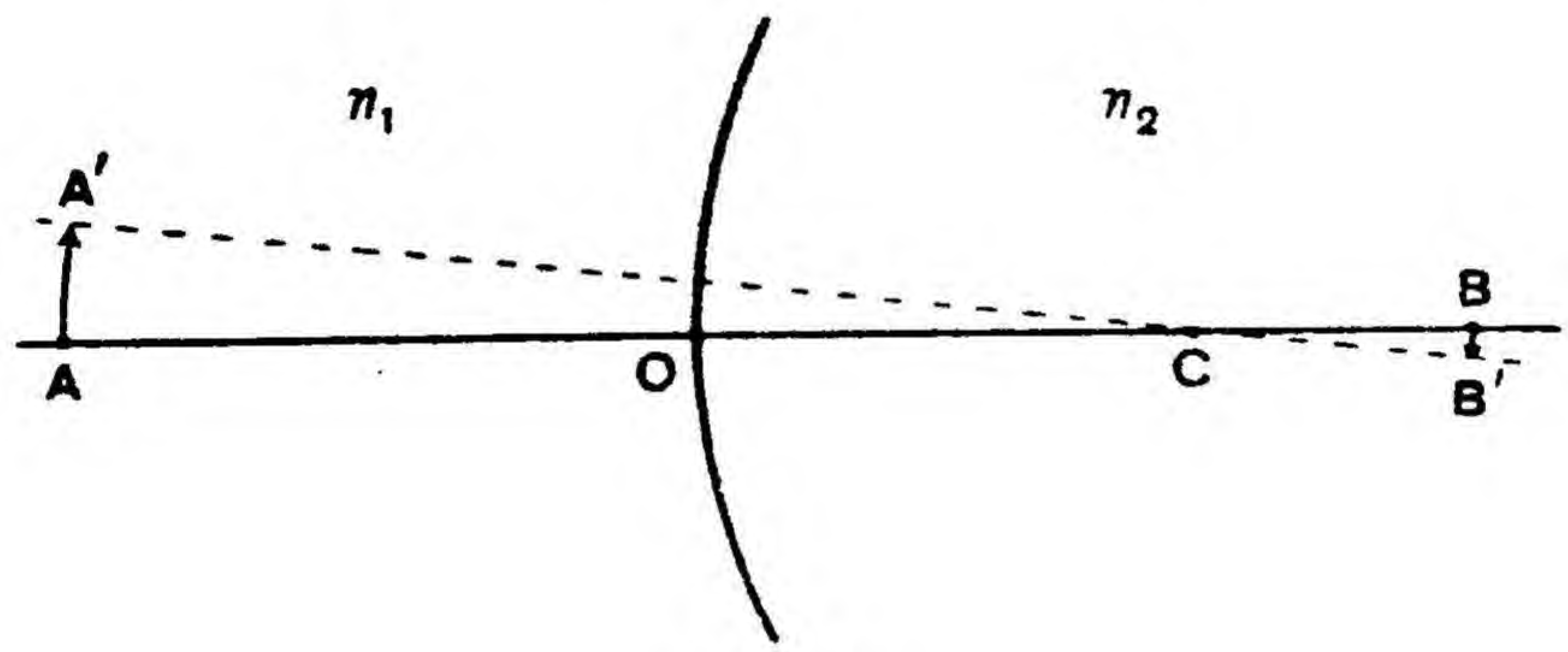


Fig. 16.

angle about the centre of curvature C of the surface; the point object A becomes a small arc of a circle AA′ and the point image B an arc BB′, and it is evident that B′ is the image of A′. For, if $u=-AO$, $v=-BO$, and $r=+CO$ satisfy the equation $\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{r}$, then so must $u=-A'O'$, $v=-B'O'$, $r=+CO'$ since $AO=A'O'$, $BO=B'O'$, and $CO=CO'$. It follows at once that the arc BB′ is the image of the arc AA′. If the angle of rotation is small, as it must be since we are restricted to paraxial rays, then AA′ and BB′ are straight lines normal to the axis of the surface. We can now state our third theorem.

Theorem 3. A spherical refracting surface forms a line image normal to the axis of a small line object normal and close to the axis and it also forms a plane image normal to the axis of a small plane object normal and close to the axis of the system. An important corollary of this arises when A′ is at infinity, when the rays from it form a parallel beam inclined at a small angle to the axis. They do not come to a focus at the second principal focus F_2, but in a plane through that point perpendicular to the axis, called the second focal

plane. They come to a focus at the point F_2' where the ray in the beam passing through C and therefore undeviated by the refraction intersects the plane (Fig. 17).

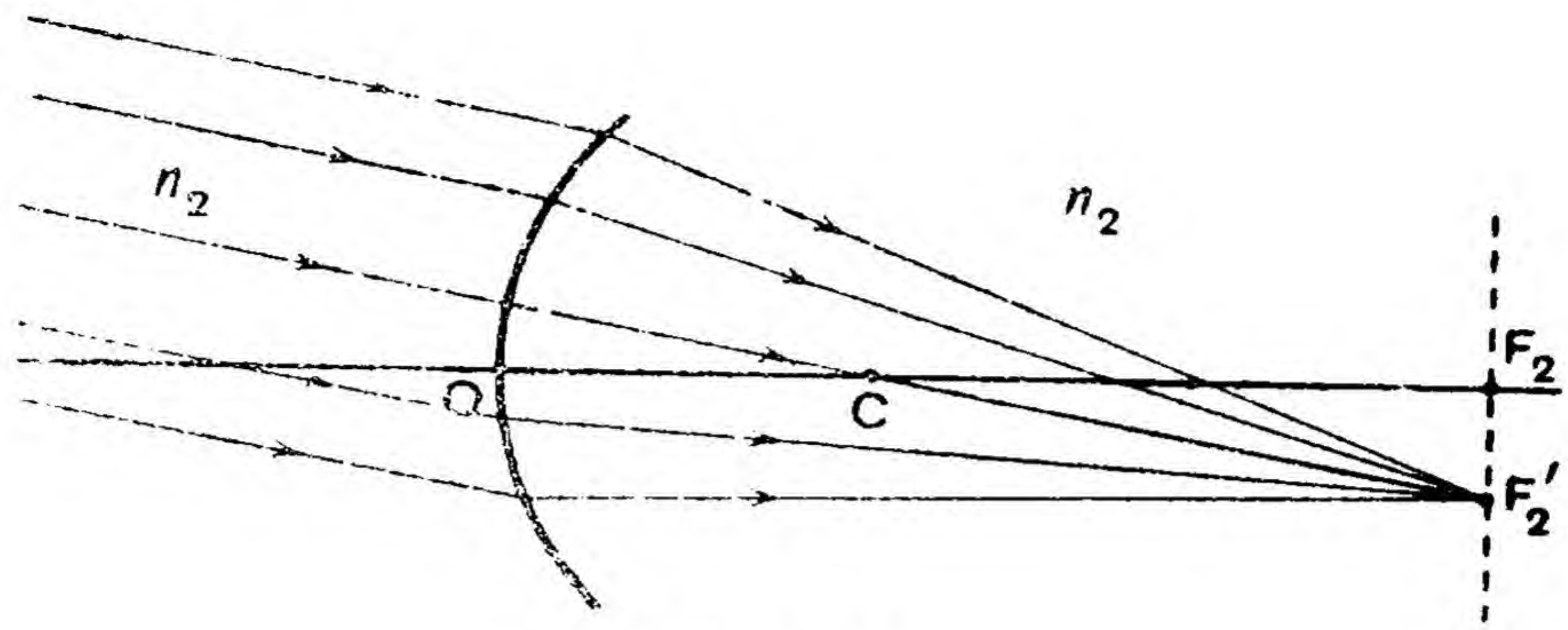


Fig. 17.

13. THE THIN LENS

We now proceed to apply the results we have proved for a single spherical refracting surface to the analysis of the properties of a lens, which consists of a piece of refracting material, such as glass, whose surfaces are portions of spheres. Most lenses are used in air, but for generality we shall assume that we are working in a medium of refractive index n_1, that of the material of the lens being n_2, $n_2 > n_1$ (Fig. 18). Let

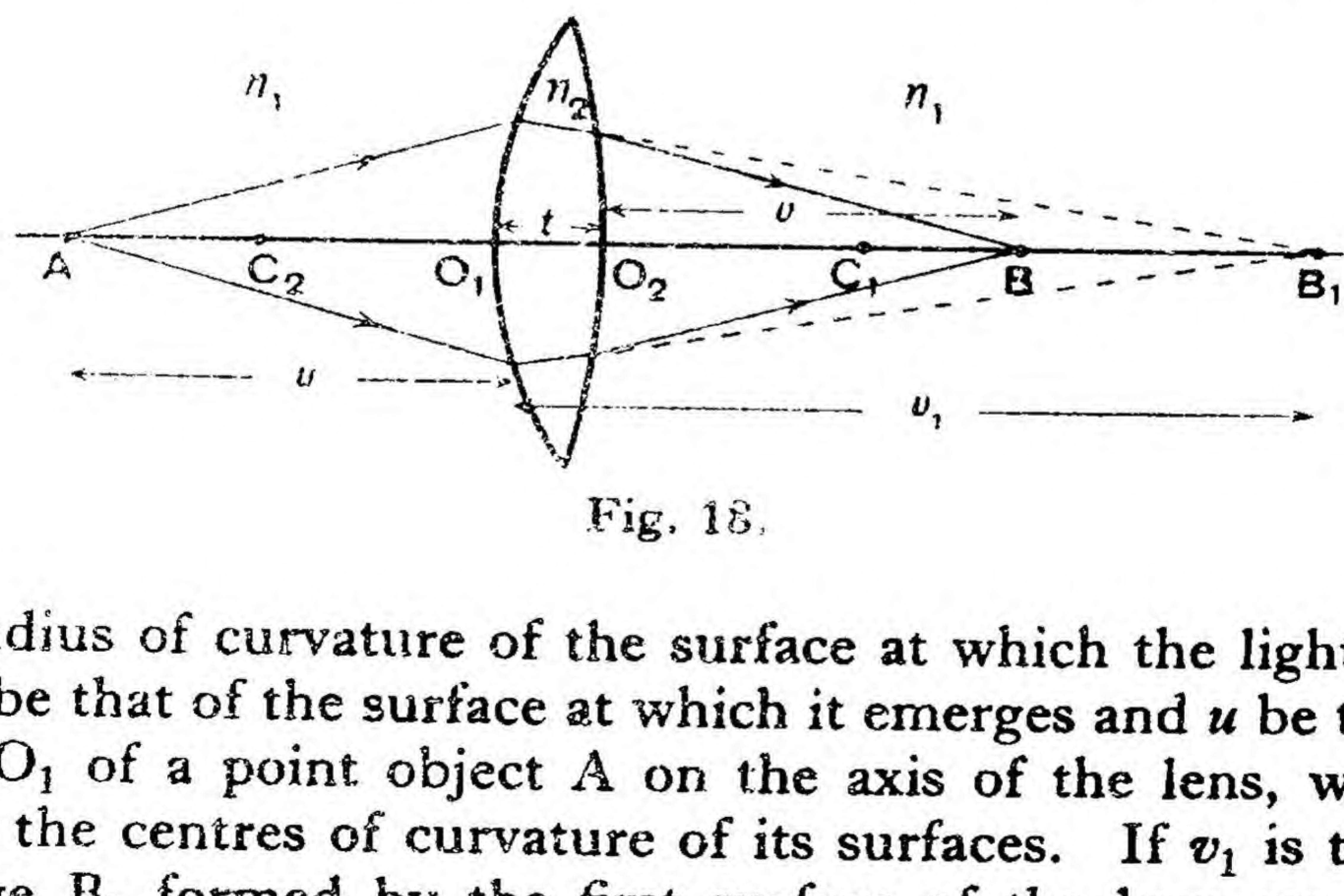


Fig. 18.

r_1 be the radius of curvature of the surface at which the light enters the lens and r_2 be that of the surface at which it emerges and u be the position relative to O_1 of a point object A on the axis of the lens, which is the line joining the centres of curvature of its surfaces. If v_1 is the position of the image B_1 formed by the first surface of the lens, we have from equation (6)

$$\frac{n_2}{v_1} - \frac{n_1}{u} = \frac{n_2 - n_1}{r_1} \qquad \text{(8)}$$

If v is the position relative to O_2 of the image B of the virtual object B_1 formed by the second surface, we have

$$\frac{n_1}{v} - \frac{n_2}{v_1 - t} = \frac{n_1 - n_2}{r_2} \qquad \text{(9)}$$

So we see that there is no simple relation between the position of the image and object in the case of a lens of finite thickness. But *if t* is negligible, equation (9) becomes

$$\frac{n_1}{v}-\frac{n_2}{v_1}=\frac{n_1-n_2}{r_1}$$

Adding this to equation (8), we have

$$\frac{n_1}{v}-\frac{n_1}{u}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad . \quad (10)$$

This relation, which is only true for paraxial rays, enables us to calculate the position of the image of a given object formed by a given **thin lens,** as a lens of negligible thickness is called. It should be noticed that we have now imposed yet another limitation, namely, that the distance between the two refracting surfaces of the lens is negligible. This so-called thin lens has quite simple properties, but it is an abstract lens, which cannot be realised in practice. It remains to be seen how close actual lenses approximate to it and how useful it will be in interpreting their properties.

From equation (10) we see that, if $u=\infty$

$$\frac{n_1}{v}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

So we have the first important property of a thin lens, which we shall call theorem 4.

Theorem 4. Rays parallel to the axis of a lens all pass through the same point on its axis after emerging from the lens. This point is called its **second principal focus** F_2 and **its position relative to the centre of the lens is the focal length,** f. The centre of the lens is the point at which the axis of the lens intersects the lens itself; it is the centre of its circular outline.

$$\therefore \quad \frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

In the same way, the first principal focus of the lens F_1 is a point on its axis such that rays from it emerge from the lens parallel to its axis. Its position relative to the centre of the lens is denoted by f_1 and so

$$\frac{n_1}{\infty}-\frac{n_1}{f_1}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

$$\therefore \quad \frac{n_1}{f_1}=-(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

We see that $f=-f_1$; this is because the media on the two sides of the lens are the same. Since the second principal focus of a lens is the more important of the two and the more frequently used, we shall usually call

it the focus for the sake of brevity. We shall only use its full name when both foci are being referred to.

$$\therefore \quad \frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

From equations (10) and (11) we have

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

which is the fundamental equation satisfied by a thin lens.

The reader should realise that, although Fig. 18 is drawn for a convex lens, the equations leading to this final relation have all been written in

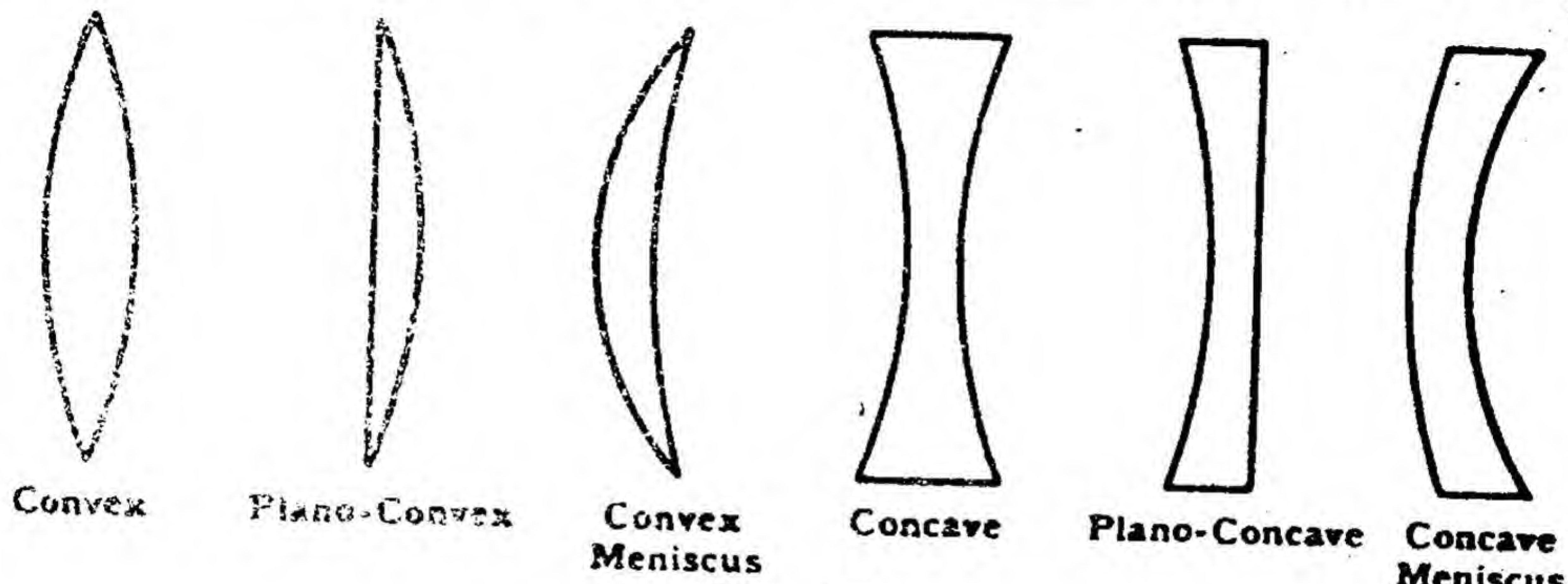


Fig. 19.

terms of positions and are therefore valid for any type of lens producing any type of image, and so this final relation is true for all cases of the six

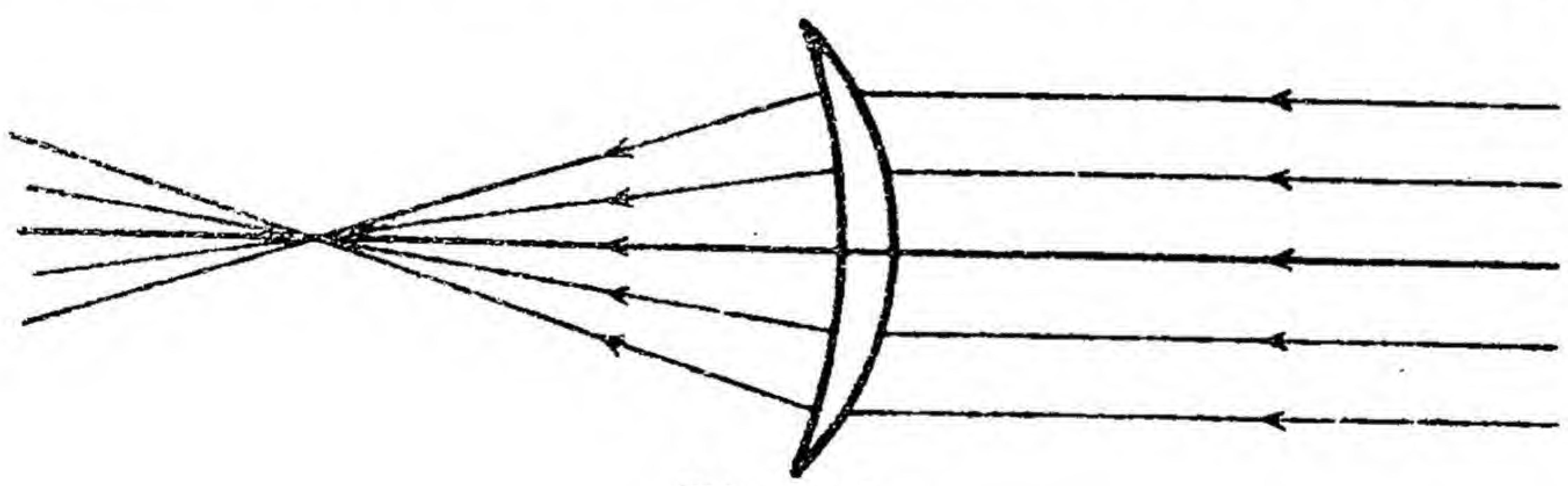


Fig. 20.

types of thin lens illustrated in Fig. 19. He should verify for himself from equation (11) that the first three types, convex, plano-convex, and

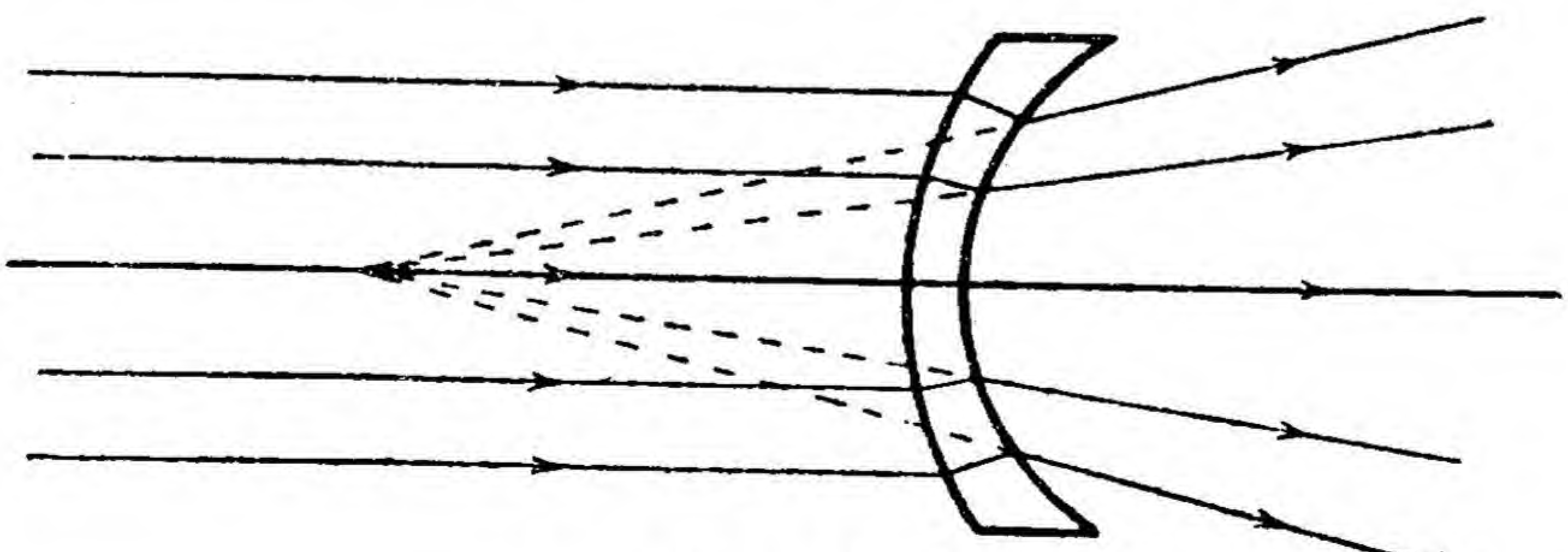


Fig. 21.

convex meniscus, are **converging lenses**; that is, they change a beam of rays parallel to the axis into a converging beam (Fig. 20). Whereas

the remaining three types, concave, plano-concave, and concave meniscus, are **diverging lenses**; that is, they change a beam of rays parallel to the axis into a diverging beam (Fig. 21). The focal length of a converging lens is always positive, while that of a diverging lens is negative; this convention agrees with that used by practical opticians, such as spectacle makers and lens designers.

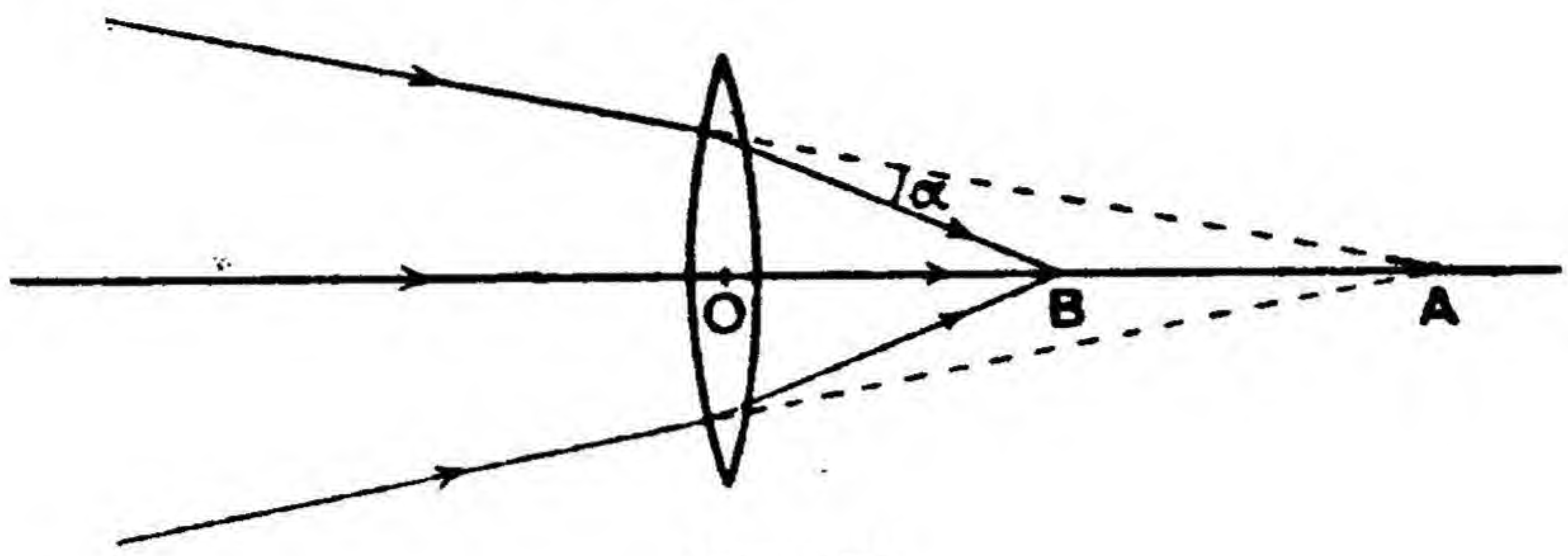


Fig. 22.

Let the rays be converging towards a virtual object at A on the axis of a converging lens (Fig. 22) and let B be the real image of this virtual object. The positions of A and B are related by equation (12). Let A tend to O; then $u \to o$; it follows from equation (12) that $v \to o$. Therefore $(u-v) \to o$ and $\alpha \to o$. That is, when the virtual object is at the centre of the lens, so is its real image and the rays forming it are undeviated (Fig. 23). So we have the second important property of a thin lens.

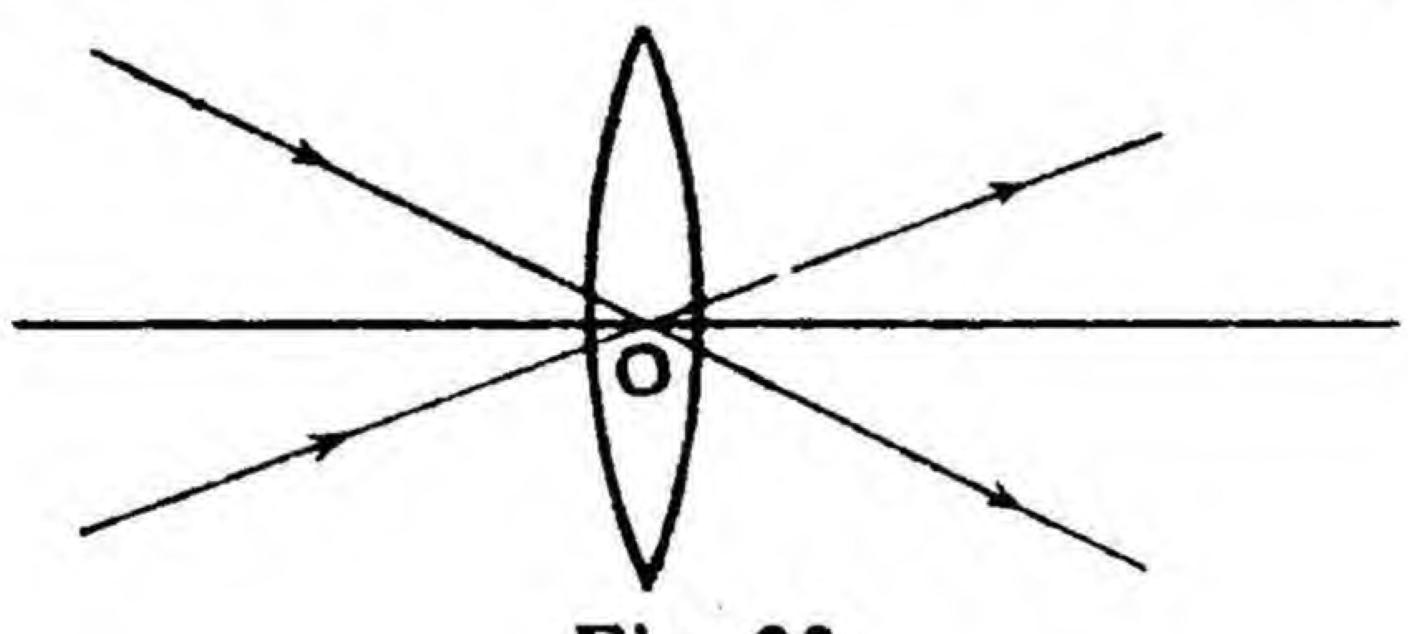


Fig. 23.

Theorem 5. All paraxial rays passing through the centre of a thin lens are undeviated.

Since a thin lens is just two co-axial single refracting surfaces acting in succession, it follows from Theorem 3 that such a lens forms a sharp plane image normal to the axis of small plane objects normal and close

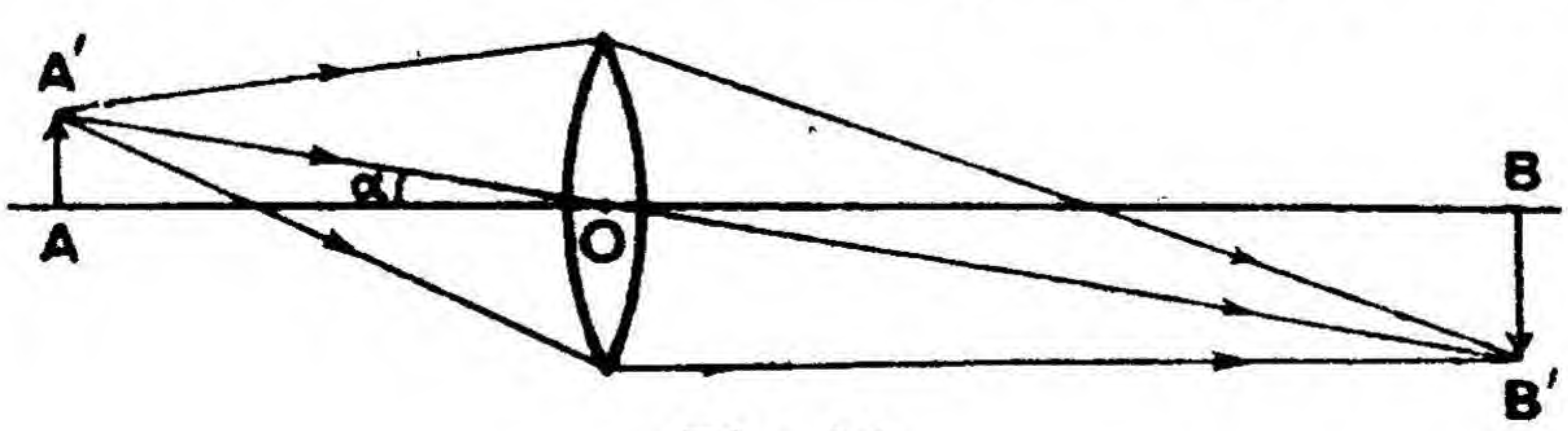


Fig. 24.

to the axis. So if AA′ is a small line object normal to the axis of a thin lens (Fig. 24), its image BB′ is found by calculating the position of B from equation (12), B′ being at such a distance vertically below B that A′OB′ is a straight line. If AA′ tends to infinity while the $\angle A\hat{O}A'$ or α

remains constant, the beam of rays from A′ tends to a parallel beam at an angle α to the axis of the lens. But as AA′ tends to infinity, BB′ tends to FF′ (Fig. 25), a line in the **focal plane** of the lens, **which is a plane through its focus and normal to its axis.** F′ is the point of intersection of the ray passing undeviated through O and the focal plane. So a parallel beam of rays inclined at a small angle to the axis of a thin lens comes to a focus at that point in the focal plane of the lens where the ray in the beam passing through the centre of the lens intersects the focal plane. It follows from the law of the reversibility of rays of light

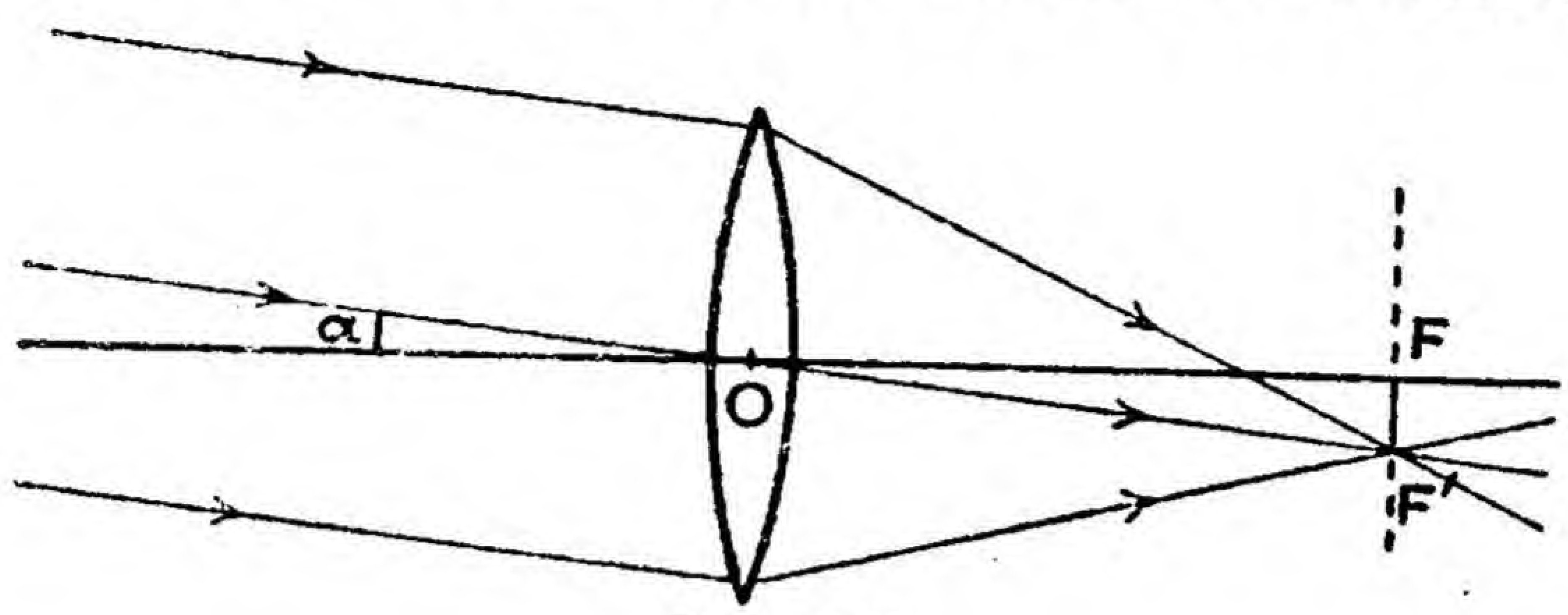


Fig. 25.

that the rays from a point object in the first focal plane of a lens emerge as a parallel beam inclined at the same angle to the axis as the ray from the point object going through the centre of the lens undeviated.

So far, then, we have deduced the following results for the effect of our abstract thin lens on paraxial rays :

(1) There are two kinds of lens, converging and diverging.

(2) Rays parallel to the axis of a lens pass through or appear to diverge from a point on its axis called the focus of the lens.

(3) **The position of the focus relative to the centre of the lens is called its focal length** and depends on the nature of the material of the lens and its form.

(4) All rays passing through the centre of a lens are not deviated.

(5) A lens forms a sharp plane image of small plane objects normal and close to its axis.

14. THE PROPERTIES OF THIN LENSES

We shall now deduce a few simple properties of thin lenses, using both the fundamental equation and a simple graphical method.

Converging lens : Case I : A real object further from the lens than twice its focal length.

If $f=+d$, and $u=-a$, we have from equation (12)

$$\frac{1}{v}-\frac{1}{-a}=\frac{1}{+d}$$

$$\therefore \quad \frac{1}{v}=\frac{1}{d}-\frac{1}{a}$$

Since $a>2d$, $+d<v<+2d$, so a real image is produced lying between the focus and a point twice the focal length from the lens.

We can verify this result and obtain the other properties of the image by constructing the image. Since we have already *proved* that a thin lens does form a line image perpendicular to the axis of a line object close and perpendicular to the axis, we only need to find where two rays from one point on the object cross after emerging from the lens in order to locate the image of the line. If AA′ is the object, draw a ray A′D parallel to the axis of the lens, which passes through the focus F after emerging from the lens (Fig. 26); then draw the ray A′O passing through the centre of the

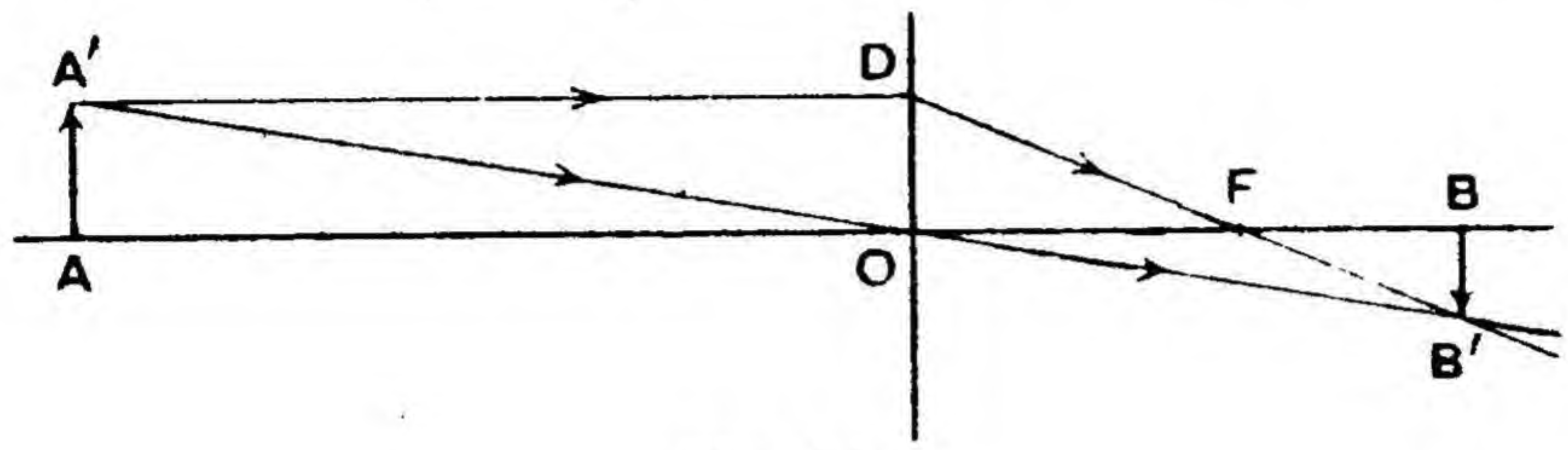


Fig. 26.

lens which goes straight on. These two rays intersect at B′, which is therefore the image of A′ and if B′B is drawn at right angles to the axis of the lens to cut it at B, then BB′ is the image of AA′. It is an interesting exercise for the reader to prove the thin lens formula $\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$ from this figure; it follows at once from the fact that the two pairs of triangles AOA′, BOB′ and DOF, B′BF are similar. The diagram shows that the image is real, inverted, and diminished. It also gives us the magnification of a thin lens, which is defined as $\frac{\text{the length of the image}}{\text{the length of the object}}$. So the magnification, m, is given by

$$m=\frac{BB'}{AA'}=\frac{OB}{OA}$$

If m is treated as a quantity having sign as well as magnitude and is defined as positive when the image is erect, we have

$$m=\frac{v}{u} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

The results of the remaining cases of interest for both converging and diverging lenses are given in Table 1, and the reader should verify them both from the equation and by drawing. In particular, he should satisfy himself that the above equation gives the correct value not only for the magnitude but also for the sign of the magnification in each case. Fig. 27 is the drawing illustrating the real image formed by a diverging lens of a virtual object inside its focus. The reader should notice the symmetry

exhibited by the results in this table, as shown, for example, by the similarity of the effect of a *converging* lens on a *virtual* object in any position to that of a *diverging* lens on a *real* object in any position.

We may conclude this deduction of the properties of the thin lens from the law of refraction with a reminder that, although the diagrams have

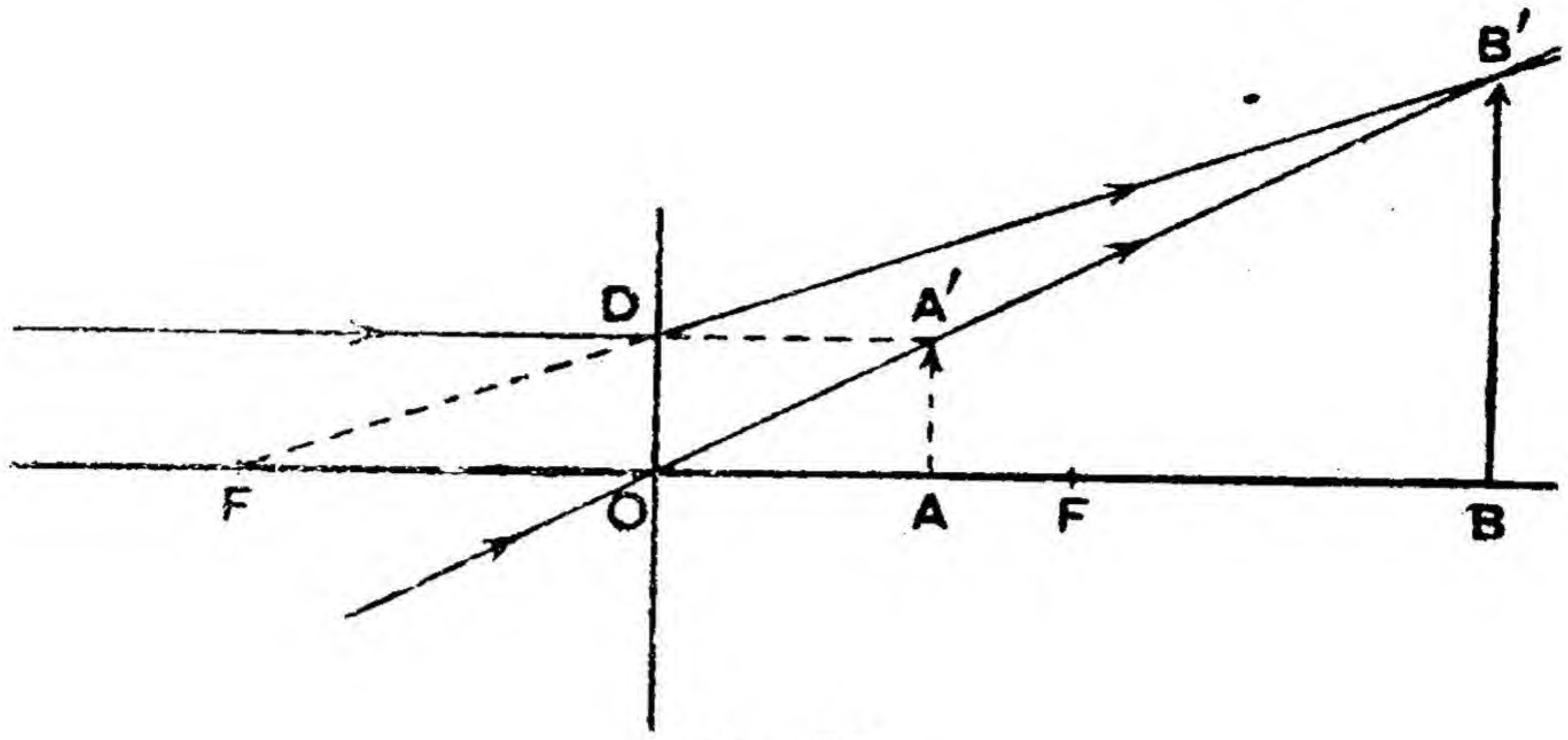


Fig. 27.

been drawn for accuracy and clearness with rays making quite large angles with the axis of the lens or refracting system, the position of the images found is only true for *paraxial* rays. Indeed we do not yet know

TABLE 1.

Nature of Lens.	Nature and Position of Object.	Nature and Position of Image.
Converging.	Real and outside $2f$.	Real, inverted, diminished, and between f and $2f$.
Converging.	Real and at $2f$.	Real, inverted, the same size as the object, and at $2f$.
Converging.	Real and between $2f$ and f.	Real, inverted, magnified, and outside $2f$.
Converging.	Real and inside the focus.	Virtual, erect, magnified, and further from the lens than the object.
Converging.	Virtual in any position.	Real, erect, diminished, and inside the focus.
Diverging.	Real in any position.	Virtual, erect, diminished, and inside the focus.
Diverging.	Virtual outside the focus.	Virtual, inverted, and outside the focus.
Diverging.	Virtual inside the focus.	Real, erect, magnified, and further from the lens than the object.

if a sharp plane image of small plane objects near the axis will be formed at all if a wide pencil of rays passes through the system ; since a sharp image is formed under paraxial restrictions it is probable that it will *not* be formed when those restrictions are removed, but we will investigate this point later.

15. EXPERIMENTAL VERIFICATION OF THE THEOREMS CONCERNING A THIN LENS

The theorems concerning a thin lens are all logical deductions from the law of refraction expressed by equation (1), which is based on experimental fact, therefore the theorems *must* be true, unless there is a flaw in our arguments somewhere. There is no question of any hypothesis or guess here; we have adhered to strict logic throughout. So the reader may wonder what point there is in comparing our results with experience. The point is this: our theorems are true only for the idealised thin lens, which cannot be realised in practice. They will only be useful in interpreting the behaviour of optical instruments and thereby improving them, if actual lenses do show some close approximation to the thin lens, and that is the point we are really going to test. We shall not describe the tests fully as they will be familiar to most students and they can be found in the elementary text-books on Light. We will take first of all the simple properties of the thin lens summarised at the end of Art. 13. If a parallel beam of rays is produced in a smoke-box and a convex lens is placed in its path with its axis parallel to the beam, it is found that all the rays do pass through a point on the axis after emerging from the lens. The same result is true of a plano-convex lens and a convex meniscus, thus proving our classification of these three lenses as converging to be correct. If the focus of the rays is examined, it will be found that it is not a geometrical point but a small patch of finite area and that the area of this patch decreases if the lens is stopped down by an iris diaphragm. This result suggests that the patch of finite area is due to the rays striking the edge of the lens having a finite angle of incidence; that is, they do not satisfy paraxial conditions. But the more nearly these conditions are satisfied, the more nearly does the result agree with our prediction. It can be shown with the same apparatus that rays passing through the centre of any lens are undeviated, if they do not make too large an angle with the axis. This result is only approximate, of course, since the rays usually produced in a smoke-box have a finite width, which would mask any small deviation. Finally, it is well known that lenses of all classes do produce sharp plane images of small plane objects near to the axis, as can easily be proved by casting the image of a small illuminated gauze on a screen with a lens. It is well to have a fairly large gauze and then it can be seen that the outer portions of the image are not in focus when the central portion is, which is what we expect from our results, which only apply to *small* objects. With this same arrangement, we can also verify all the properties of thin lenses set out in Table 1. In fact, some of those cases have very common practical applications. The real inverted diminished image produced by a converging lens of a distant object is used in the camera; the real, inverted, magnified image produced by a converging lens of an object just outside its focus is used

in the projection lens of a cinematograph or epidiascope. But it is obvious already that a single lens would be no use in these cases, as the edges of the images would be badly out of focus, since the rays to them are by no means paraxial, so we shall have to investigate the case of rays at a finite angle to the axis of the lens. The virtual, erect, magnified image produced by a converging lens of an object inside its focus is exemplified in the simple magnifying glass and the virtual, erect, diminished image of any object produced by a diverging lens is familiar to short-sighted persons who wear spectacles made of concave lenses. So we see that our theorems are all verified by experiments of a qualitative kind, and the reader will carry out for himself experiments to verify the quantitative relation which is true for all lenses

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

and he will be able to see for himself to what extent this is satisfied by actual lenses. Table 2 is a typical set of results obtained in an experiment of this kind, and the surprising thing is that actual lenses, whose thickness

TABLE 2

A LENS EXPERIMENT

Diameter of lens : 4·0 cm. ; thickness 0·3 cm.

u cm.	v cm.	$\frac{1}{u}$ cm.$^{-1}$	$\frac{1}{v}$ cm.$^{-1}$	$\frac{1}{u}+\frac{1}{v}$ cm.$^{-1}$	f cm.
26·0	88·8	0·3846	0·1126	0·4972	
29·0	66·2	0·3448	0·1511	0·4959	
32·0	54·2	0·3125	0·1845	0·4970	
35·0	47·6	0·2857	0·2101	0·4958	
38·0	43·1	0·2632	0·2320	0·4952	
40·3	40·3	0·2481	0·2481	0·4962	
			Average	0·4962 ±0·0007	20·15 cm.

can by no stretch of imagination be called negligible, do satisfy it to an accuracy of 1 in 700. So our abstract thin lens is going to be a very useful tool in the analysis of actual optical instruments, and we shall illustrate its use in connection with the projection lantern.

16. THE CALCULATION OF THE FOCAL LENGTH OF THE LENS OF A PROJECTION LANTERN

The purpose of a projection lantern is to throw a magnified image of a lantern slide on to a screen some distance from the lantern. The image must be large enough to be visible from all parts of the room and its size therefore depends on the size of the room itself. It is evident from our theory of the thin lens that a converging lens will be required for the

purpose, and we shall now proceed to calculate the focal length of the lens suitable for a particular case. Let us suppose that the lantern is to be used in a room some 30 ft. square and that it is required to put the lantern so that the slide is 25 ft. from the screen, and that a picture 6 ft. square is to be produced from a slide of standard size, $3\frac{1}{4}$ in. square. Using the usual notation we have the following two simultaneous equations in a and b, which are the *distances* of the object and image respectively from the lens :

$$a+b=300$$

$$\frac{b}{a}=\frac{72}{3\cdot25}$$

whence $b=287$ and $a=13\cdot0$.

Now for a thin lens
$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

and remembering that u and v represent **positions** we have in this case $u=-13\cdot0$ and $v=+287$. Substituting these values in the above equation, we get

$$f=+12\cdot4 \text{ in.}$$

From the equation

$$\frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

we can calculate values of r_1 and r_2 when we have settled the material of which the lens is to be made and so we can send in the specification of the lens we require to the firm which is to make the lantern. We know that they cannot make us a thin lens, but we know that the lens they will make is so near in its behaviour to the above thin lens that it will do its work satisfactorily. The reader will notice that the values of r_1 and r_2 are not uniquely specified ; that is, there are an infinite number of pairs of values of r_1 and r_2 which will satisfy our requirements, and so this leaves us some latitude if any further demands should be made upon us when we come to consider rays making finite angles with the axis or the elimination of colour effects.

We may also need to use the lantern at a distance of 10 ft. from the screen to enable the lecturer to operate it himself. The reader should work out the focal length of the lens needed to give the same size picture as before. He will find that it comes to $5\cdot0$ in. We see that we need a more powerful lens to produce the same magnification in a smaller distance, which is just what we should expect. This simple application of the theory of the thin lens to a practical problem in the choice of lenses for lantern or cinematograph projection is a good example of the scientific method of approaching a problem Before the theory of lenses was known, this problem could only be solved by the method of trial

and error, which would consist in setting up a lantern in the room and trying lenses of various focal lengths until the right ones had been found. The advantage of the scientific method is obvious; the focal length of the required lens is calculated from the well-established theory of lenses. It is true—and this is nearly always the case—that our theoretical lens cannot quite be reproduced in practice, but the manufacturer can make something approximating so closely to it that our prediction is of real use to him. We shall return to a consideration of this problem after we have considered the defects of the image produced by a single lens.

17. THE LENS AS A SET OF PRISMS

It is instructive to consider another way of looking at the lens, which is illustrated in Fig. 28, in which a ray of light enters a convex lens at E and leaves it at F. If BE and CF, the tangent planes to the surface of the lens at E and F respectively, meet at A the ray behaves just as if it was passing through the prism ABC. A moment's consideration will show that the greater the distance from the axis at which the ray passes through the lens, the greater the angle α of this equivalent prism. So a convex lens may be replaced by a set of prisms with refracting edges all perpendicular to the axis of the lens and of gradually increasing angle as we go outwards from the axis. We shall now establish the fundamental equation for paraxial rays passing through a thin lens on this basis.

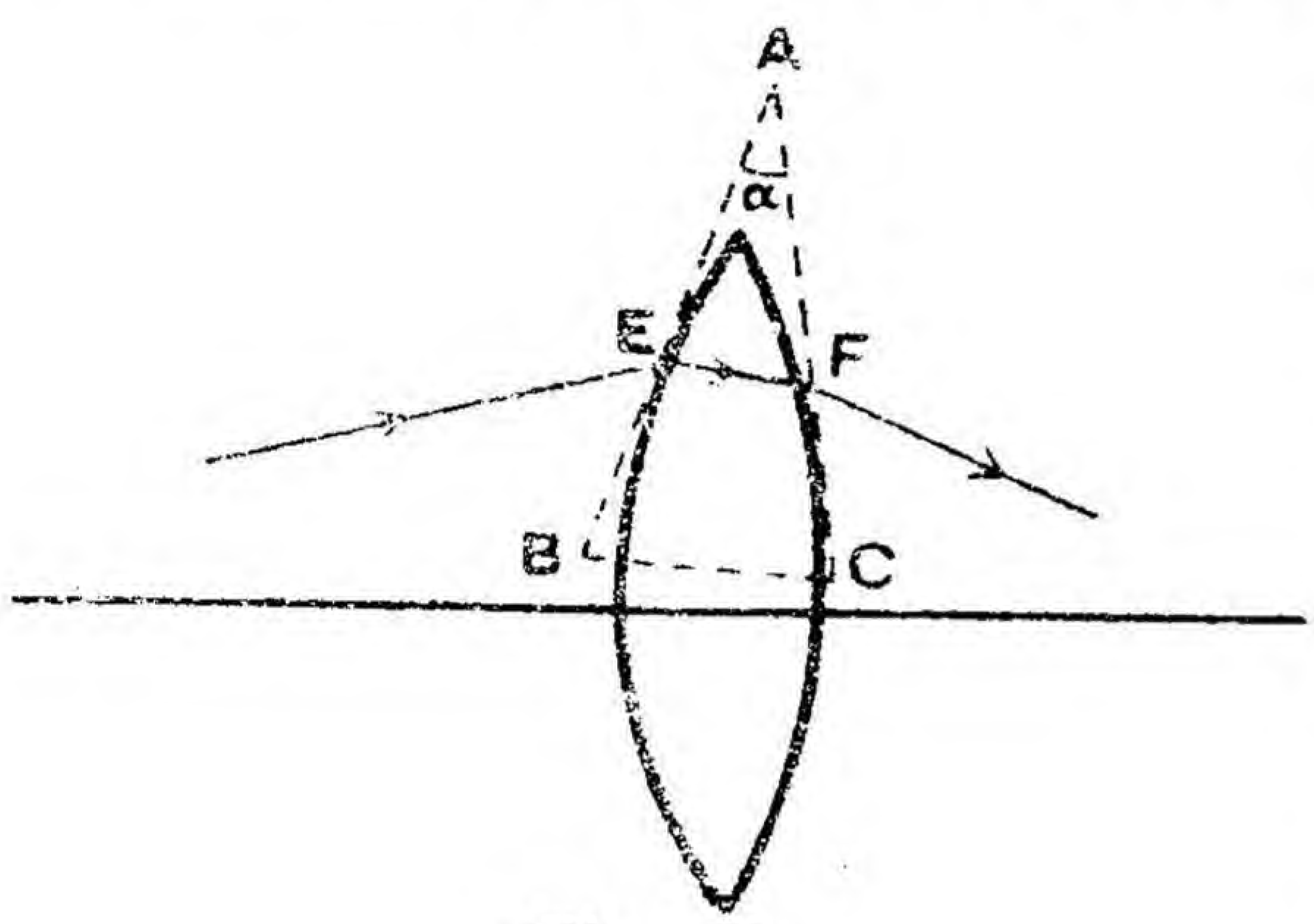


Fig. 28.

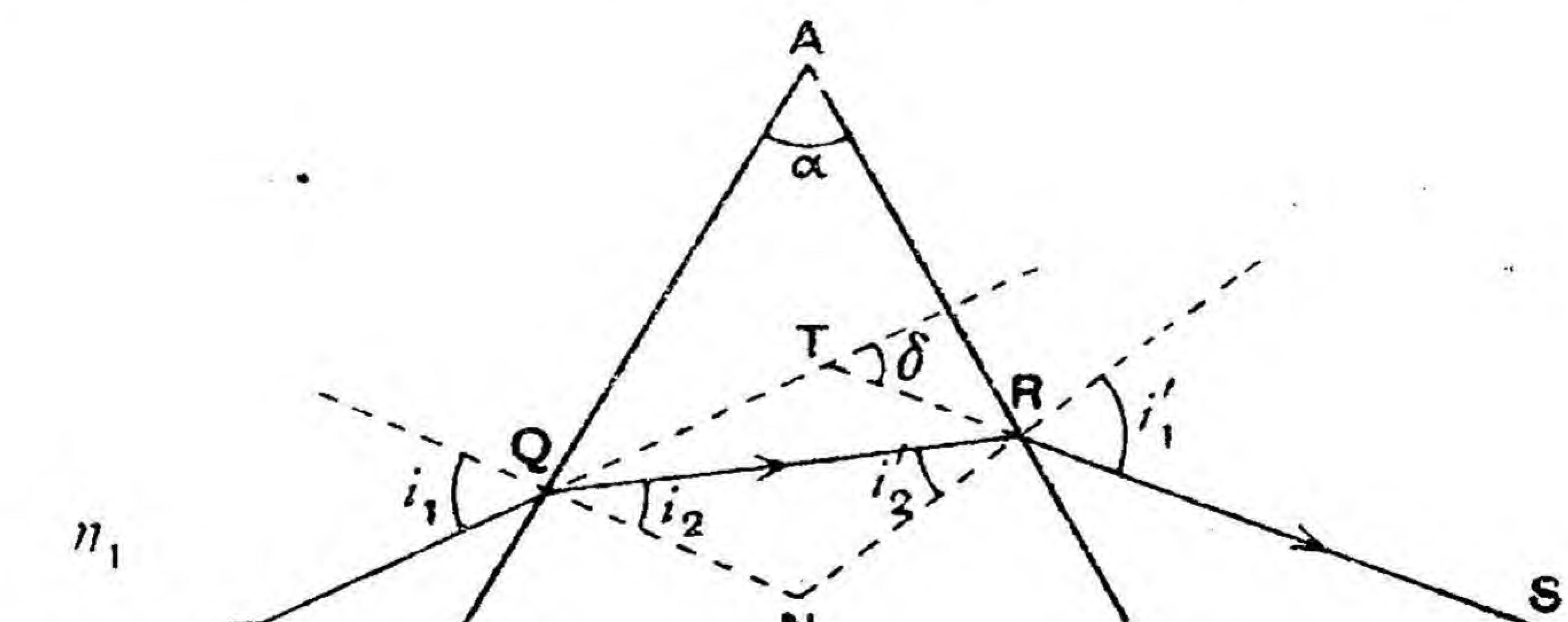


Fig. 29.

We shall first find the deviation of a ray of light passing at nearly normal incidence through a prism of small angle, since these conditions

are satisfied for paraxial rays. If PQRS is a ray passing through the prism ABC of angle α producing a deviation δ, QN and RN being the normals to AB and AC respectively, we have (Fig. 29)

$$\begin{aligned}\delta &= \angle T\hat{Q}R + \angle T\hat{R}Q \\ &= (i_1 - i_2) + (i_1' - i_2') \\ &= (i_1 + i_1') - (i_2 + i_2')\end{aligned}$$

If the incidence is nearly normal, all these angles are small and so the law of refraction may be written

$$n_1 i_1 = n_2 i_2, \; n_1 i_1' = n_2 i_2'$$

$$\therefore \qquad \delta = \frac{n_2}{n_1}(i_2 + i_2') - (i_2 + i_2')$$

$$\therefore \qquad \delta = \left(\frac{n_2}{n_1} - 1\right)(i_2 + i_2')$$

But $i_2 + i_2' = \pi - \angle QNR = \alpha$, since AQNR is a cyclic quadrilateral.

$$\therefore \qquad \delta = \left(\frac{n_2}{n_1} - 1\right)\alpha \qquad \text{. . . . . (14)}$$

independent of the angle of incidence.

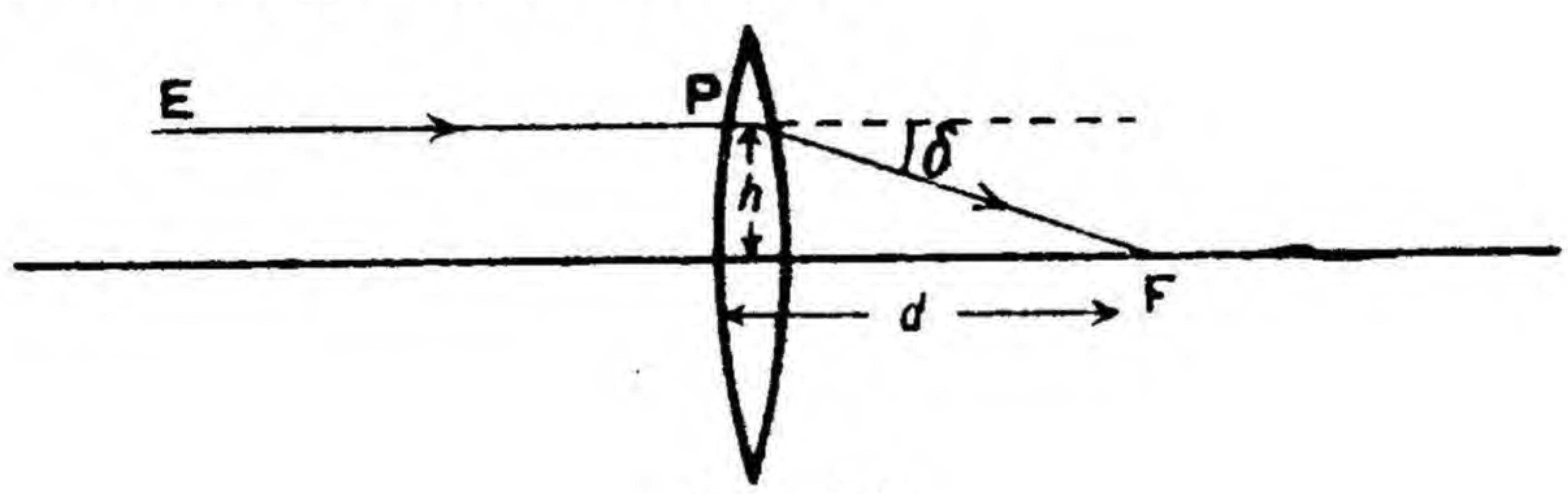


Fig. 30.

Consider a ray EP parallel to the axis of a lens at a distance h from it and let it cut the axis at F after emergence (Fig. 30). Then the deviation, δ, produced by the lens is given by

$$\delta = \frac{h}{d}$$

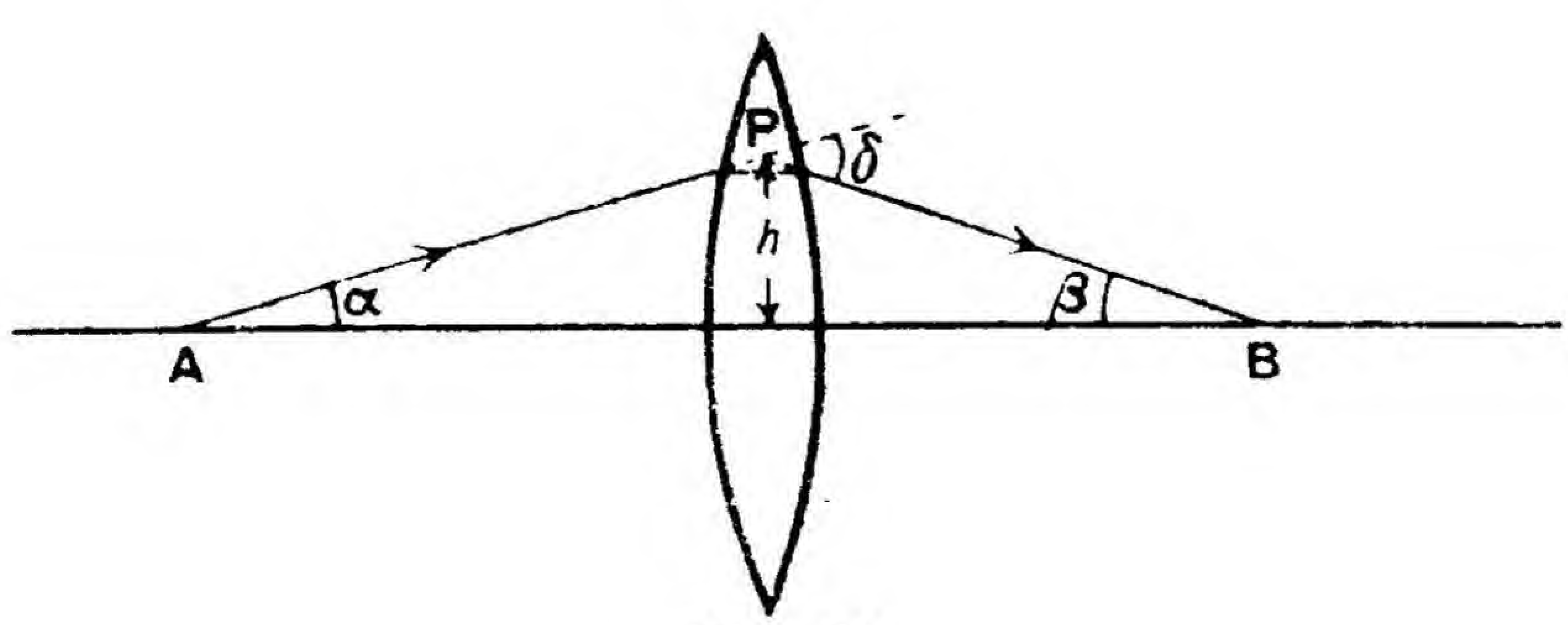


Fig. 31.

Now consider a ray AP (Fig. 31) making a small angle α with the axis ; it must also be deviated through an angle δ, since it strikes the lens at the same distance h from the axis as the ray EP and the deviation produced

by the prism equivalent to this zone of the lens is independent of the angle of incidence. Let the emergent ray PB cut the axis at B. Then we have

$$\delta=\alpha+\beta$$

Since all the angles are small, as we are dealing only with paraxial rays,

$$\frac{h}{d}=\frac{h}{a}+\frac{h}{b}$$

$$\therefore \quad \frac{1}{b}+\frac{1}{a}=\frac{1}{d} \quad \ldots\ldots\ldots \quad (15)$$

We must now show that this equation is true for any value of h. If α_1 is the angle made by the first face of the equivalent prism at P with a plane normal to the axis of the lens, then α_1 is equal to the angle between the normal to the first face and the axis of the lens.

$$\therefore \quad \alpha_1=\frac{h}{c_1},$$

where c_1=the numerical value of the radius of curvature of the surface at which the light enters the lens. If α_2 is the corresponding angle for the second face of the equivalent prism, then

$$\alpha_2=\frac{h}{c_2}$$

where c_2=the numerical value of the radius of curvature of the surface at which the light leaves the lens.

Therefore α, the angle of the equivalent prism, is given by

$$\alpha=\alpha_1+\alpha_2=h\left(\frac{1}{c_1}+\frac{1}{c_2}\right)$$

$$\therefore \quad \delta=\left(\frac{n_2}{n_1}-1\right)\alpha=\left(\frac{n_2}{n_1}-1\right)h\left(\frac{1}{c_1}+\frac{1}{c_2}\right)$$

also

$$\frac{1}{d}=\frac{\delta}{h}$$

$$\therefore \quad \frac{1}{d}=\left(\frac{n_2}{n_1}-1\right)\left(\frac{1}{c_1}+\frac{1}{c_2}\right)$$

and is independent of h. This equation together with equation (15) shows that all paraxial rays from A pass through B after emerging from the lens; in other words, B is the image of A. Finally, if we express our results in positions instead of distances, we have

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

and

$$\frac{1}{f}=\left(\frac{n_2}{n_1}-1\right)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

the standard equations for a lens.

The value of this method of looking at a lens is that it derives the standard relation from the point of view of the lens as a device for bending rays either towards or away from the axis and from it emerges the valuable property that the bending produced by a given zone of the lens is the same whatever the angle at which the incident ray strikes it, provided only that it is small. The reader will find it interesting to derive from this constant deviation how the image will move as the object moves from infinity up to and beyond the centre of the lens. It can be applied both to converging and diverging lenses.

18. THE POWER OF A LENS

We have seen that the real function of a lens is to bend all rays inwards or outwards, and so the more powerful a lens the greater the angle through which it will bend a ray. Since the deviation is proportional to the distance of the ray from the axis of the lens, the rational definition of the power of a lens is the deviation it produces in rays at unit distance from its centre.

Now $$\delta=\frac{h}{f}$$

$$\therefore \quad \text{Power}=\frac{\delta}{h}=\frac{1}{f}=\text{P, say.}$$

So **the power of a lens is the reciprocal of its focal length** and the unit of power is the **dioptre**, which is the power of a lens of focal length 1 metre. So the power of a lens in dioptres is given by the reciprocal of its focal length expressed in metres.

The power of a single refracting surface is defined in a similar way, and is the ratio of the refractive index of the second medium to the second focal length of the surface. It follows from equation (6) that

$$\text{P}=\frac{n_2}{f_2}=\frac{n_2-n_1}{r}$$

So the equation for refraction at a single spherical surface

$$\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{r}$$

may be re-written in the form

$$\frac{n_2}{v}-\frac{n_1}{u}=\text{P}$$

For a thin lens of refractive index n in air we have by putting $n_2=n$ and $n_1=1$ in equation (11)

$$\frac{1}{f}=(n-1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

Now $\dfrac{n-1}{r_1}=P_1$, the power of the first surface

and $\dfrac{1-n}{r_2}==P_2$, the power of the second surface

∴ If P is the power of the thin lens,

$$P=P_1+P_2$$

Finally, if we have two thin lenses of focal length f_1 and f_2 in contact, the reader can easily prove that the focal length, f, of the combination is given by

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2}$$

This is most simply done by considering a point source on the axis at infinity as the object for the first lens, which produces a real image of it at its focus and that is treated as a virtual object for the second lens The result can easily be extended to any number of lenses in contact, when we have

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}+\quad . \quad . \quad .$$

where f is the focal length of the combination and f_1, f_2, f_3 . . . are the focal lengths of the individual lenses.

$$\therefore \quad P=P_1+P_2+P_3+\quad . \quad . \quad .,$$

the mathematical expression of theorem 6.

Theorem 6: The power of a number of thin lenses in contact is equal to the sum of the powers of the component lenses.

19. THE THIN LENS BY FERMAT'S PRINCIPLE

We have seen that the two fundamental laws of rays of light can be expressed in Fermat's Principle of Stationary Time (Art. 4), and we shall use it to prove the standard relation for a thin lens. If A is a point object

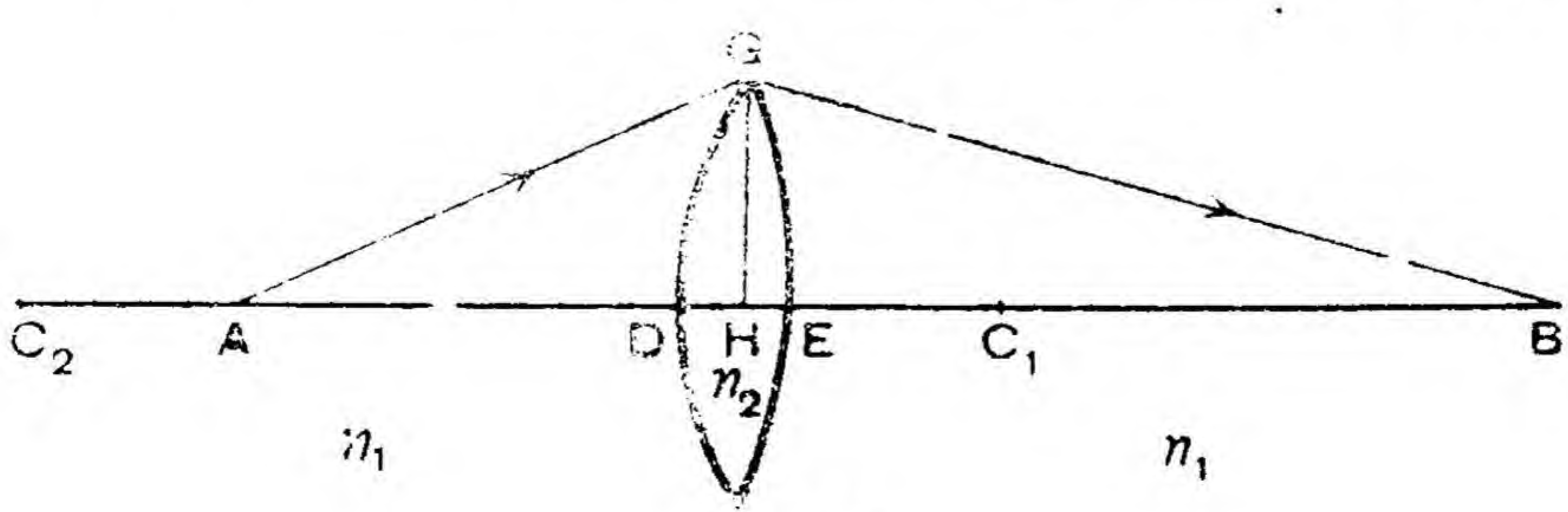


Fig. 32.

on the axis of the lens and B is its point image, let us consider the two rays ADEB and AGB going from A to B (Fig. 32). By Fermat's principle, all the rays take the minimum (or maximum) time and so any two rays

must take the *same* time. If c_1 and c_2 are the velocities of light in the media of refractive index n_1 and n_2 respectively, we have

$$\frac{AG+GB}{c_1}=\frac{AD+EB}{c_1}+\frac{DE}{c_2}$$

But $$\frac{c_1}{c_2}=\frac{n_2}{n_1}$$

by Fermat's principle applied to the general law of refraction.

$$\therefore \quad n_1(AG+GB)=n_1(AD+EB)+n_2DE$$

$$\therefore \quad n_1\{(AH^2+GH^2)^{\frac{1}{2}}+(BH^2+GH^2)^{\frac{1}{2}}\}$$
$$=n_1(AH-HD+BH-HE)+n_2DE$$

If we restrict ourselves to paraxial rays and thin lenses, GH is small compared to AH or HB. So we may expand by the binomial theorem and neglect terms in $\left(\frac{GH}{AH}\right)^4$ or $\left(\frac{GH}{BH}\right)^4$ and higher powers.

$$\therefore \quad n_1\left\{AH\left(1+\frac{GH^2}{AH^2}\right)^{\frac{1}{2}}+BH\left(1+\frac{GH^2}{BH^2}\right)^{\frac{1}{2}}\right\}$$
$$=n_1(AH+BH)+(n_2-n_1)(DH+HE)$$

$$\therefore \quad n_1\left\{AH\left(1+\frac{GH^2}{2AH^2}+\,.\,.\right)+BH\left(1+\frac{GH^2}{2BH^2}+\,.\,.\right)\right\}$$
$$=n_1(AH+BH)+(n_2-n_1)(DH+HE)$$

$$\therefore \quad n_1\left\{AH+\frac{GH^2}{2AH}+BH+\frac{GH^2}{2BH}\right\}$$
$$=n_1(AH+BH)+(n_2-n_1)(DH+HE)$$

$$\therefore \quad \frac{n_1}{AH}+\frac{n_1}{BH}=\frac{2(n_2-n_1)(DH+HE)}{GH^2}$$

If C_1 is the centre of curvature of the surface of the lens at which the light enters and C_2 that at which it leaves, then

$$DH(2DC_1-DH)=GH^2$$

As we are dealing with a thin lens, DH can be neglected compared to DC_1;

$$\therefore \quad 2DH.DC_1=GH^2$$

$$\therefore \quad \frac{2DH}{GH^2}=\frac{1}{DC_1}$$

Similarly $$\frac{2HE}{GH^2}=\frac{1}{EC_2}$$

$$\therefore \quad \frac{n_1}{AH}+\frac{n_1}{BH}=(n_2-n_1)\left(\frac{1}{DC_1}+\frac{1}{EC_2}\right)$$

or expressing the equation in positions instead of distances,

$$\frac{n_1}{v}-\frac{n_1}{u}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

from which the equations

$$\frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

and

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

follow at once. So we see that Fermat's Principle of Stationary Time also leads to the usual equation relating the positions of the object, image, and focus of a thin lens.

20. ASTIGMATIC LENSES

So far we have confined ourselves exclusively to lenses whose surfaces are spherical and therefore symmetrical about the axis of the lens. If the incident rays diverge from a point on the axis, the emergent rays are also symmetrical about the axis and come to a point focus. We shall conclude this account of the thin lens with a brief discussion of astigmatic lenses, in which one or both faces are cylindrical and which are therefore not symmetrical about the axis of the lens. Consequently a beam from a point on the axis does not emerge symmetrical relative to the axis and so does not come to a point focus, which is the reason for the name astigmatic, from the Greek α, without, and $\sigma\tau\iota\gamma\mu\alpha$, a point.

Let us consider in the first place a lens both of whose faces are cylindrical with the axes parallel to each other and let it be mounted so that the axes are vertical. It follows at once that for rays in any horizontal plane the focal length of the lens is given by the usual relation

$$\frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

where r_1 and r_2 are the radii of curvature of the faces at which the light enters and leaves respectively, n_2 is the refractive index of the material of the lens and n_1 that of the medium in which it is immersed, $n_2>n_1$. Also the focal length for rays in any vertical plane is infinity, since for such rays the lens behaves like a plane parallel-sided plate. If, for the sake of simplicity, we consider a convex lens of this type, then it will be a converging lens for rays in any horizontal plane. If the object is a point on the axis of the lens at infinity, the incident pencil consists of rays parallel to the axis of the lens. This will be bent inwards in a horizontal direction but will be undeviated in a vertical direction. So the emergent pencil is wedge shaped instead of conical, and it converges to a vertical **line focus** at a distance f from the lens. It is therefore an astigmatic pencil.

Now let us consider the rather more complicated case of a lens in which one face is spherical of radius of curvature r_1 and the other cylindrical of radius of curvature r_2 and let the lens be mounted so that the axis of the cylindrical face is vertical. Then the lens has two focal lengths, one $f_\perp$ for rays in a plane perpendicular to the cylindrical axis given by

$$\frac{n_1}{f_\perp}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

and the other f_{11} for rays in a plane parallel to the cylindrical axis and given by

$$\frac{n_1}{f_{11}}=\left(\frac{n_2-n_1}{r_1}\right)$$

If for the sake of argument we consider a lens of this type in which both faces are convex, then it will be a converging lens for rays in either of the above planes, but the rays will be bent inwards more rapidly in a horizontal than in a vertical plane (Figs. 33 & 34). What will be the effect

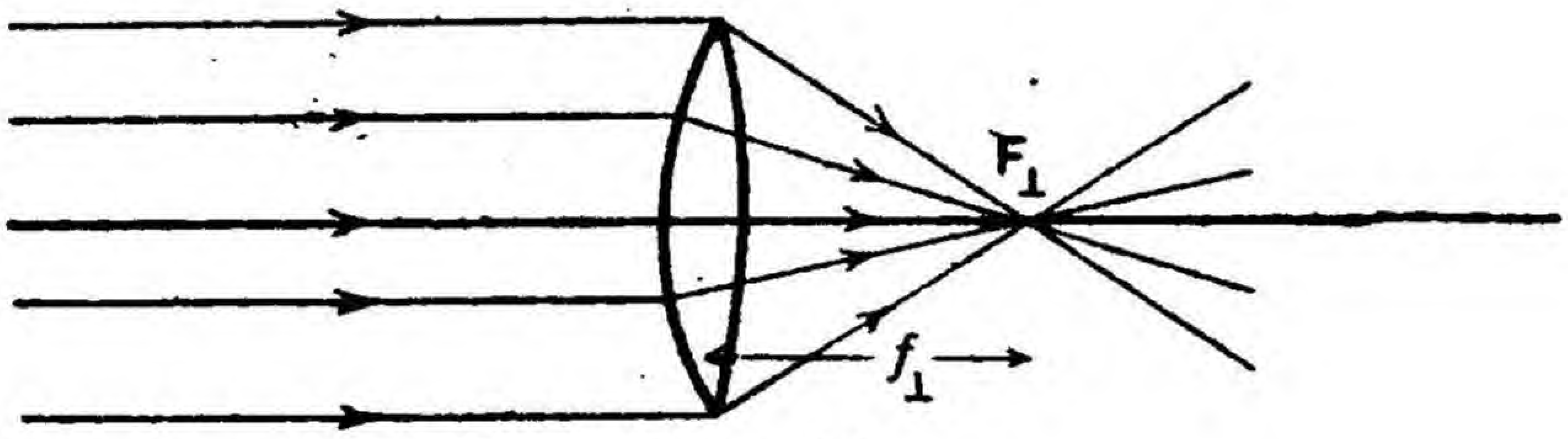


Fig. 33.

of this lens on a point object at infinity on the axis of the lens ? Since the rays are bent inwards horizontally more rapidly than vertically, the beam will come to a vertical line focus first at $F_\perp$ and then to a horizontal line focus at F_{11}. The emergent pencil is therefore an astigmatic beam,

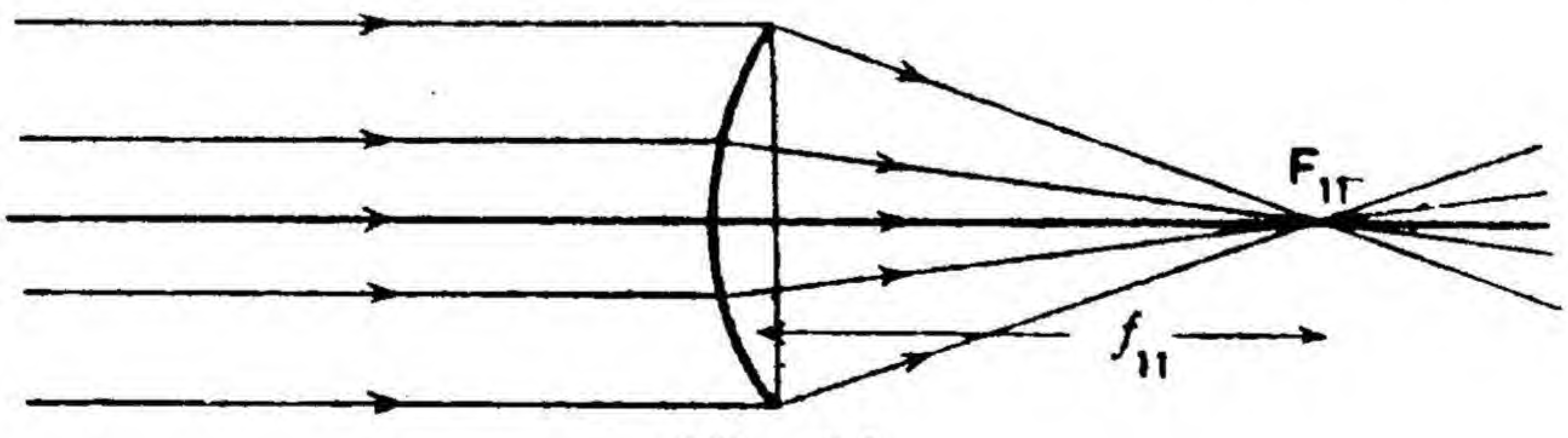


Fig. 34.

possessing two **line foci**, the nearest approach to a point focus being in a plane normal to the axis between $F_\perp$ and F_{11}, on which the **circle of least confusion** is formed. The greater the difference between the focal lengths of an astigmatic lens, the greater the distance between the two focal lines and the larger the circle of least confusion.

Since the astigmatic lens forms the focal lines for a point object at infinity on its axis, it will do the same thing for a point object on its axis at a finite distance away. The arguments which were used in the case of the spherical lens can also be applied here to show that the astigmatic lens will form focal lines for point objects close to its axis. Therefore

an astigmatic lens will form a slightly blurred plane image of a small plane object perpendicular and close to the axis, if the image is received on a screen at the position of the circle of least confusion. This is the best focus which can be obtained from an asymmetrical lens of this sort, which has two different focal lengths in two mutually perpendicular planes. It is also interesting to consider the formation by such a lens of an image of a gauze placed with its wires parallel and perpendicular to the axis of the cylindrical surface of the lens, which is still supposed to be vertical. If the image is received on a screen placed to receive the vertical focal line, then all the vertical lines of the gauze will be sharply in focus while the horizontal lines will be quite out of focus. The reason for this is that each point of a vertical wire forms a small vertical line and these various lines overlap to form a single sharp vertical line rather longer than would be produced by a spherical lens forming an image of the same gauze on the same screen. In the same way each point of a horizontal wire forms a small vertical line and these lines combine to form a faint oblong patch on the screen with its long side horizontal (Fig. 35). If the screen is now moved back so as to receive the horizontal focal line, the vertical wires will go out of focus while the horizontal ones will be sharply in focus (Fig. 35).

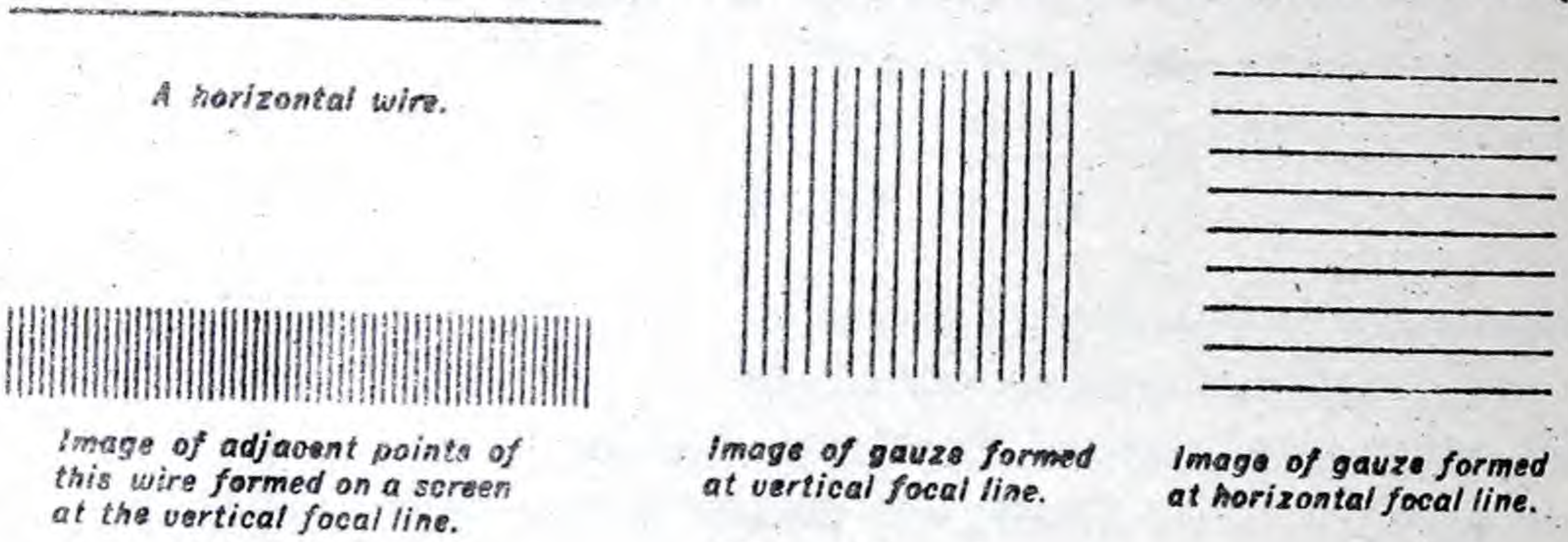


Fig. 35.

Somewhere in between the circle of least confusion will be obtained, when a slightly blurred image of the vertical and horizontal wires will be produced, the best image of the whole gauze which can be obtained. The lack of symmetry of an astigmatic lens is beautifully exemplified in its property of being able to focus either the vertical wires alone or the horizontal ones, but not both together. It is clear that even this property is lost if the gauze is not placed with its wires parallel and perpendicular to the axis of the cylindrical surface of the lens.

Astigmatism is a very common fault of the human eye, which often has slightly different focal lengths in two planes which are not always at right angles to each other. This defect results in a slight loss of sharpness of focus; things cannot be seen as clearly as with the normal eye; it will also happen that a person suffering from astigmatism will be able to adjust his eyes so as to focus only the vertical strips of mortar in a brick wall or only the horizontal ones but not both sets at once. This defect can be remedied by the use of suitable cylindrical lenses, as

will be shown later on when dealing with the eye more fully, but enough has already been said to make it clear that, if one of the lenses of a pair of round spectacles correcting for astigmatism falls out of its frame, it cannot be replaced anyhow, as there is only one position of the lens in the frame in which the correction will be properly made.

What have we achieved so far? From the two " axioms " for rays of light we have proved by logical deduction the following theorems for paraxial rays:

(1) Any number of co-axial spherical refracting surfaces forms a sharp plane image perpendicular to the axis of small plane objects perpendicular and near to the axis.

(2) the equation $\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$ is true for a thin lens of any material immersed in any medium.

(3) The power of a set of thin lenses in contact is equal to the sum of the powers of the component lenses.

(4) These theorems are very useful for the interpretation of the behaviour of actual lenses, for it is found that ordinary lenses obey them to an accuracy of less than one per cent. in spite of the fact that they have a finite thickness and paraxial conditions are not always satisfied.

Our next problems are quite clear; we must apply our results to reflection, which is a particular case of refraction; we must then investigate the effect of lenses (and mirrors) on rays making a finite angle with the axis.

EXAMPLES ON CHAPTER II

1. Find an expression relating the position of an object to the position of its image as formed in a single spherical refracting surface.

A distant object is focused on the ground glass screen of a camera when the screen is at a distance of 10 cm. from the lens. If the camera is filled with water of refractive index 1·3 what will be the new distance of the screen from the lens when the object is in focus, assuming that the face of the lens nearest to the screen is convex and has a radius of curvature of 20 cm.? (*Camb. Schol.*)

2. A refracting system consists of a block of glass with a convex spherical surface placed in air. Find the position of the principal foci if the radius of the surface be r and the refractive index n. (If formulæ are used they must be proved.)

The power of a system is defined to be the angle subtended at the pole by a distant object divided by the length of the image. From this definition show that the power is $(n-1)/r$ in this case. (*London Inter.*)

3. The eye can be regarded as a single spherical refracting surface, the cornea, of radius of curvature 7·8 mm., separating media of refractive index 1·00 and 1·34. Find at what distance from the pole of the surface a beam of parallel rays will come to a focus.

Suppose the eye is now directed at a point object 150 mm. away, what radius of curvature will the cornea have to assume in order to focus the rays from this object at the same point as the above parallel beam? Also what radius of curvature

will the cornea have to assume if the eye is to focus a parallel beam at the same point as in the above two cases when the eye is under water (refractive index 1·30) ?

4. A camera has an equiconvex lens of 10 cm. focal length, the refractive index being 1·55. How far must the plate be from the lens to photograph distant objects under water (refractive index 1·3) when the camera is itself under water and is filled with (a) air, (b) water ? Would it make any difference to the result if the distant objects were in air instead of under water ?

5. A point image is formed of a point object on the axis of a set of co-axial spherical refracting surfaces. Prove that a displacement of the object in a given direction along the axis produces a displacement of the image in the same direction.

6. State and prove the relation connecting the position of the object and the image of a lens.

The image of an object placed 30 cm. from a lens is formed at a point 15 cm. on the opposite side of the lens. What is the type of the lens and what is its focal length ? (*Oxford Schol.*)

7. A convex lens of focal length 30 cm. produces a real image which is three times the size of the object. At what distances from the lens are the object and image situated ? (*Oxford Schol.*)

8. When a pin is placed 10 cm. in front of a lens and a plane mirror 20 cm. behind the lens it is found that an erect image of the pin is formed in the plane of the pin itself. Draw a picture showing the rays by which an eye sees the image, and calculate the focal length of the lens. (*Camb. Schol.*)

9. Deduce an expression for the focal length of a lens in terms of the radii of curvature of its surfaces and the refractive index of its material.

Two narrow vertical slits are in a plane perpendicular to the axis of a converging lens. The lens produces real images of the slits 3 cm. apart. If the lens is moved 60 cm. along its axis, real images of the slits $\frac{1}{3}$ cm. apart are again produced in the same plane as before. Find the distance apart of the slits and the focal length of the lens. (*Camb. Schol.*)

10. Find an expression for the focal length of a thin lens in terms of the radii of curvature of the faces of the lens and the refractive index of the glass.

A plano-concave water lens is formed between a thin glass plate and one surface of a thin double convex lens in contact with it. The focal length of the combination is 26·2 cm. and of the convex lens alone 19 cm. The glass lens is then turned over so that the other surface is in contact with the glass plate, and the focal length of the combination is now 25·4 cm. Find the refractive index of the glass, and the radius of curvature of each face of the lens. (Refractive index of water $=4/3$.) (*Camb. Schol.*)

11. A metal plate containing an illuminated circular hole is placed at one end of an optical bench, and a screen at the other. By means of a convex lens an image of the hole is formed on the screen, the diameter of the image being 2·25 cm. If the lens is moved 20 cm. along the bench an image of the hole again appears on the screen, its diameter being now 1·00 cm. What is the real size of the hole and how far is it from the screen ? (*O. and C.*)

12. A real object is placed perpendicular to the axis of a convex lens so that a real image of length y is formed at a distance v from the lens. Show how to obtain the length x of the object and the focal length f of the lens by plotting y against v, and find the values of x and f if y changes from 0·80 cm. to 1·80 cm. when v changes from 30 cm. to 45 cm.

Describe in detail how you would perform the experiment and verify the value of x by direct measurement of the object. (*N.U.J.B.*)

13. Describe two methods for the determination of the focal length of a concave lens.

A thin equiconvex lens is placed on a horizontal plane mirror and a pin held 20 cm. vertically above the lens coincides in position with its own image. The space between the under surface of the lens and the mirror is filled with water (refractive index 1·33) and then, to coincide with its image as before, the pin has to be raised until its distance from the lens is 27·5 cm. Find the radius of curvature of the surfaces of the lens. (*N.U.J.B.*)

14. A point source of light is moved steadily along the axis of a simple lens from a position in contact with the lens to one an infinite distance away from it. Construct a simple diagram from which it is possible to read off the position of the image corresponding to any position of the source.

If f is the focal length of the lens, p the distance of the source from one focus, and q the distance of the image from the other focus, show that $pq=-f^2$. What advantage has this formula in a focal length measurement? (*London.*)

15. Explain in detail how you would determine the focal length of a concave lens.

An equi-convex lens rests on a horizontal plane mirror and an object held above it is found to coincide with its image when 12 cm. from the lens. On filling the space between the lens and mirror with a liquid of refractive index 1·46 the distance of the object for coincidence with the image is found to be 20 cm. Calculate (a) the radius of curvature of the lens surfaces; (b) the refractive index of the glass of the lens. (*N.U.J.B.*)

16. A converging lens is placed between a pin and a plane mirror. The three are adjusted so that (a) an erect image, (b) an inverted image, of the pin is obtained coincident with the pin. Draw diagrams illustrating the formation of the image in each case, and explain how in each case the focal length of the lens may be deduced from a knowledge of the relative positions of the pin, lens, and mirror. (*Tripos, Part I.*)

17. A plane mirror is placed at a distance d behind a convex lens of focal length f. A pin is placed at a distance u in front of the lens, so that its image is seen after refraction by the lens, reflection by the mirror, and refraction by the lens again. For what values of u may a real image of the pin be formed? (*Camb. Schol.*)

18. A projection lantern for a large hall is required to project a picture, 20 ft. × 20 ft., of lantern slides 3¼ in. × 3¼ in. Find the focal length of the thin lens needed to do this if the slide is to be (a) 100 ft., (b) 40 ft. from the screen.

What specification would you give to the lens manufacturer in each case if the lenses are to be made of glass of refractive index 1·500 and are to be (a) equi-convex, (b) plano-convex?

19. A cinematograph lens is needed to produce a picture 30 ft. × 22·5 ft. of a film 1 in. × ¾ in. on a screen 180 ft. from the lens. Find how far the film must be from the lens and also the focal length of the lens. Work out the specification you will give to the lens manufacturer if the lens is to be made of glass of refractive index 1·600 and is to be plano-convex.

20. A lens is required for a 16-mm. cinema projector to produce a picture 4 ft. × 2 ft. of a film 0·5 in. × 0·25 in. on a screen 15 ft. from the film. Find the focal length of the lens required and its specification, if it is to be equi-convex of refractive index 1·55. What would be the size of the picture produced if the operator was compelled to use the projector with the lens 10 ft. from the screen? How would the brightness of the image obtained in this case compare with that of the first case?

21. Show that in the case of a thin lens (a) all rays nearly parallel to the axis falling on the lens at any given point are equally deviated on passing through, (b) the deviation produced is different for different points on the lens and proportional to the distance of the point from the axis of the lens. Hence derive the usual formula connecting the focal length of the lens, the radii of curvature of its faces and the refractive index of the material. (*Camb. Schol.*)

22. (i) Show that all rays of light proceeding from a point A on the axis of a convex lens of small aperture, and passing through the lens to a point B, also on the axis of the lens, have the same time of travel provided that

$$\frac{1}{v}+\frac{1}{u}=(\mu-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right)$$

where

u = distance from A to the lens
v = distance from B to the lens
r_1 and r_2 = radii of curvature of the two faces of the lens

(ii) Show how it follows from this fact that B is the image of a point source at A.

(iii) Explain the convention of signs which must be used to make the expression hold in the general case of a concave lens and a virtual image.

(iv) Draw diagrams showing the course of the rays of light by which the eye sees (a) a real image, (b) a virtual image in a convex lens. *(Camb. Schol.)*

23. State and explain Fermat's Principle of Least Time.

Apply the principle to establish the relation between the focal length of a thin lens and the distances from it of a point source of light and its real image. *(Tripos, Part I.)*

24. A ray of light is incident on one face of a thin double convex lens in a direction parallel to the axis, and after two internal reflections emerges from the second face. If the lens is of focal length 50 cm., and is made of glass of refractive index 1·5, how far from the centre of the lens is the point where the emerging ray cuts the axis? *(Camb. Schol.)*

25. Show that for the case of a thin lens, $xy=f^2$ where x and y are the distances of the object and image from the corresponding focal points, and f is the focal length. Show also that the magnification of the image is y/f.

Explain how this result can be applied to the experimental determination of the focal length of a lens. *(Oxford Schol.)*

26. A thin plano-convex lens is used to form an image of the sun when it is vertically overhead, on the floor of a swimming-bath 2 metres deep. Draw a curve showing how the size of the image of the sun depends upon the focal length of the lens. (Refractive index for water = 1·3, for glass 1·5.) *(Oxford Schol.)*

27. Explain what is meant by the focal length of a lens.

Two thin converging lenses of focal length 10 cm. are placed 5 cm. apart. Obtain a relationship between the position of the object and image for the combination. Can any meaning be attached to the term focal length for the lens system? *(Oxford Schol.)*

28. Two thin convex lenses of 20 cm. focal length each are set up co-axially at a distance 5 cm. apart. The image of an object 100 metres away and 5 metres high is formed by the combination.

What is the size and position of the image?

What would be the effect of filling the space between the lenses with water? *(Oxford Schol.)*

29. Two thin convergent lenses each of focal length f are placed on the same axis at a distance f apart. Find graphically or otherwise the principal foci and the focal length of the combination.

Find also the positions of the images and the linear magnifications for small objects placed on the axis and perpendicular to it at distances f and $2f$ in front of one of the lenses. *(Oxford Schol.)*

30. Show that the deviation produced by a thin lens in a ray making a small angle with the axis is h/f, where f is the focal length of the lens and h the distance from the axis where the ray strikes the lens. Hence or otherwise show that for two co-axial thin lenses in air a finite distance apart, there exist two points for which an incident ray through the first emerges in a parallel direction through the second and that if object and image distances are measured from these two points respectively the formulæ for thin lenses apply. *(Oxford Schol.)*

31. Two thin convex lenses of focal lengths f_1 and f_2 are placed on the same axis at a distance f_1+f_2 apart. An object is placed on the axis at a distance u_1 from the lens of focal length f_1, and the final image is at a distance u_2 from the lens of focal length f_2. Find the relation between u_1, u_2, f_1, and f_2. Find also (a) the linear magnification of the image, (b) the angular magnification of the image in the case where $u_1=\infty$. *(Camb. Schol.)*

32. A pencil of light diverging from a point on the axis of a thin lens is refracted at the first surface of the lens, reflected at the other surface, and then refracted

out. Show that the final image will be real except when the distance of the luminous point from the lens is less than

$$\frac{r\,r'}{2\{\mu r-(\mu-1)r'\}}$$

where r, r' are the radii of the first and second faces of the lens respectively and μ is the refractive index of the lens material. (*Camb. Schol.*)

33. If a pin is placed 10·5 cm. in front of a convex lens of focal length 20·0 cm., a *faint* real inverted image can be seen coinciding with the pin itself. How is the image formed? Use your explanation to find the radius of curvature of the surface of the lens remote from the pin. If the lens is turned round, the pin has to be put 8·8 cm. from the lens to get the same image to coincide with the pin. Calculate the radius of curvature of the other surface of the lens and the refractive index of its material.

34. A pin is placed on one side of a converging lens, which forms a real inverted image of it on the other side. When a concave lens is placed between the image and the converging lens at a distance of 9·5 cm. from the image, a new real inverted image is formed 20·0 cm. from the diverging lens. Calculate its focal length. The concave lens is now fixed to a converging lens of focal length 10·5 cm. and the focal length of the combination is found to be 25·0 cm. Calculate the focal length of the concave lens from these readings. Which of these two methods do you consider to be the more accurate way of finding the focal length of a concave lens? Give reasons for your answer.

35. An astigmatic convex lens has one face spherical of radius 20·0 cm. and the other cylindrical with the same radius, the refractive index of the glass being 1·50. It is placed with the axis of the cylindrical face vertical at a distance of 60·0 cm. from a gauze whose wires are 1 mm. apart and horizontal and vertical. Calculate the position, nature, and distance apart of the wires of the images formed by the lens.

The gauze is now turned through 45° in its own plane. Will the lens form an image of it now? If so, where will it be and what will it be like?

Chapter III

REFLECTION AT CURVED SURFACES

21. INTRODUCTORY

It is natural to begin the study of reflection at curved surfaces with spherical surfaces as was done in the case of refraction, both because the mathematical treatment is likely to be simple, and most mirrors are spherical in shape, because the surface of a sphere is the only one which can be made with accuracy and reasonable cheapness in practice. So we shall derive the theorems for spherical mirrors from the law of reflection by logical deduction just as we did for refraction, and we shall see how our results agree with the behaviour of actual mirrors.

22. REFLECTION AT A SPHERICAL SURFACE

We have already seen in Art. 3 that reflection is a particular case of refraction for which $n=-1$, if the refraction is from air to a medium of refractive index n. It follows from the general law of refraction expressed by the equation $n_1 \sin i_1 = n_2 \sin i_2$ that here reflection is given by the condition $n_2 = -n_1$ when $i_2 = -i_1$. We can therefore derive the equation for reflection at a spherical surface from that for refraction by making this substitution. The equation for refraction at one spherical surface for paraxial rays is

$$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{r} \qquad \ldots \ldots \ldots \ldots (16)$$

If $n_2 = -n_1$, this equation becomes

$$\frac{-n_1}{v} - \frac{n_1}{u} = \frac{-n_1 - n_1}{r}$$

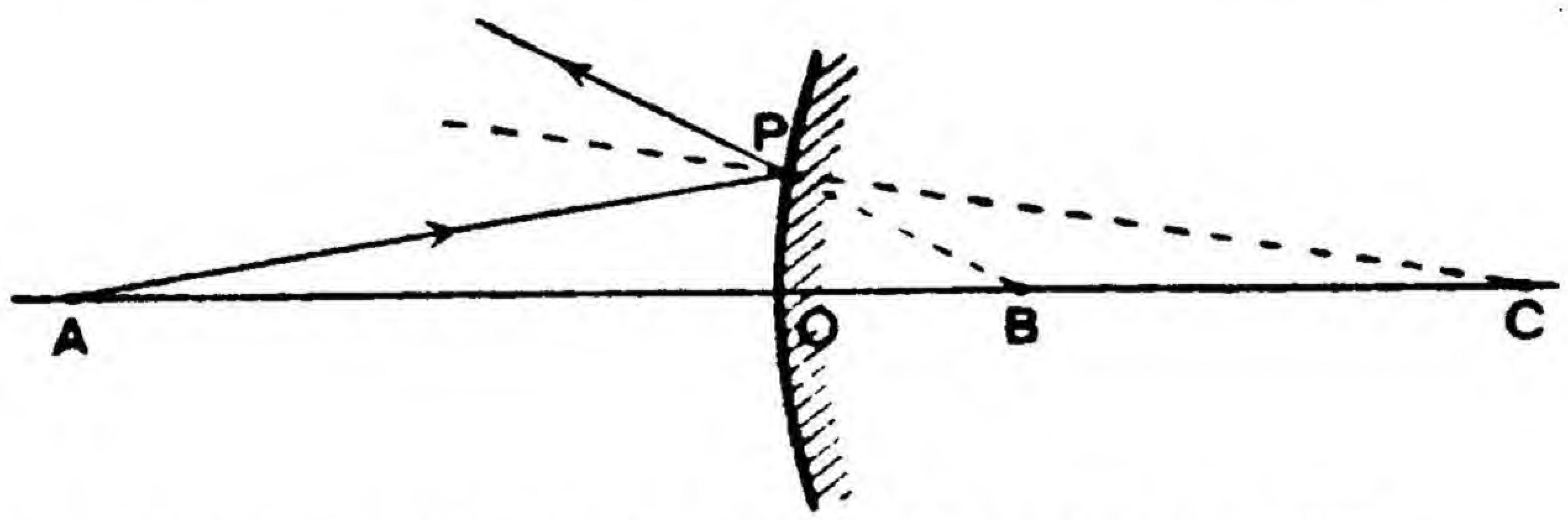
or

$$\frac{1}{v} + \frac{1}{u} = \frac{2}{r} \qquad \ldots \ldots \ldots \ldots (17)$$

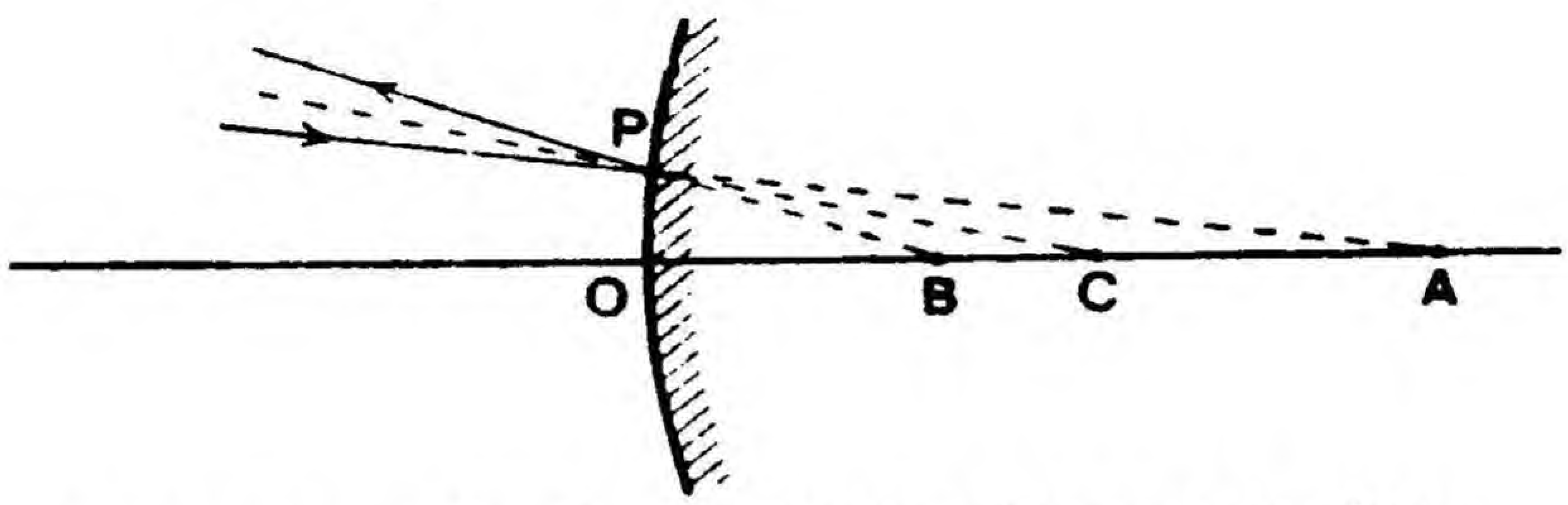
the usual equation connecting the **positions** of an object **and its image** formed by a spherical mirror.

We shall now prove this relation directly from the law **of reflection**, expressing our equations in terms of the **positions** of the object, image, and centre of curvature of the mirror. The diagrams for the six possible cases for the convex and concave mirror are shown in Fig. 36. In every case C is the centre of curvature of the mirror, O is the centre of its

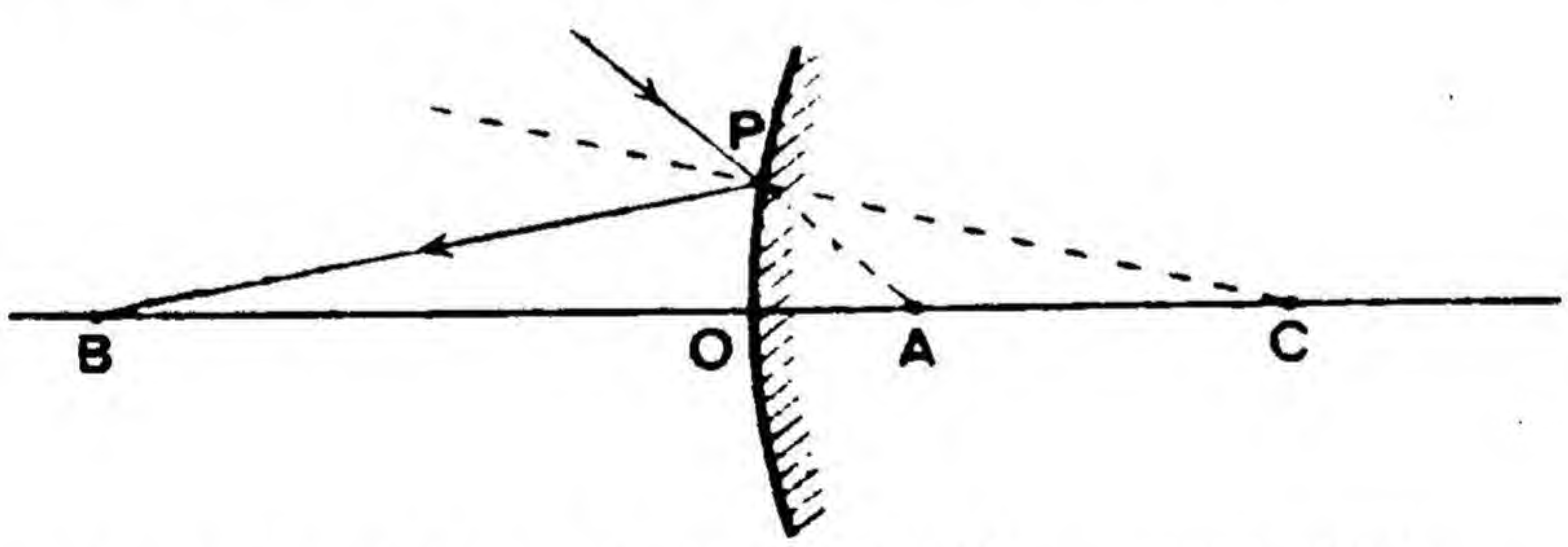
circular outline, called the pole of the mirror, AP is any ray emitted from the object A (or converging to a virtual object A) on the axis OC of the mirror; it strikes the mirror at P and is reflected along PB (or BP) to cut the axis at B (or appears to come from B after reflection). In four of the cases either the object or image is virtual. By the law of reflection,



A virtual image of a real object formed by a convex mirror



A virtual image of a virtual object formed by a convex mirror



A real image of a virtual object formed by a convex mirror

Fig. 36.

CP, produced if necessary, is the internal (or external) bisector of the angle APB.

$$\therefore \quad \frac{AP}{PB}=\frac{AC}{CB}$$

If we restrict ourselves to paraxial rays

$$\frac{AP}{PB}=\frac{AO}{BO}$$

within the limits of experimental error.

$$\therefore \quad \frac{AO}{BO}=\frac{AC}{CB}$$

$$\therefore \quad \frac{u}{v}=\frac{u-r}{r-v}$$

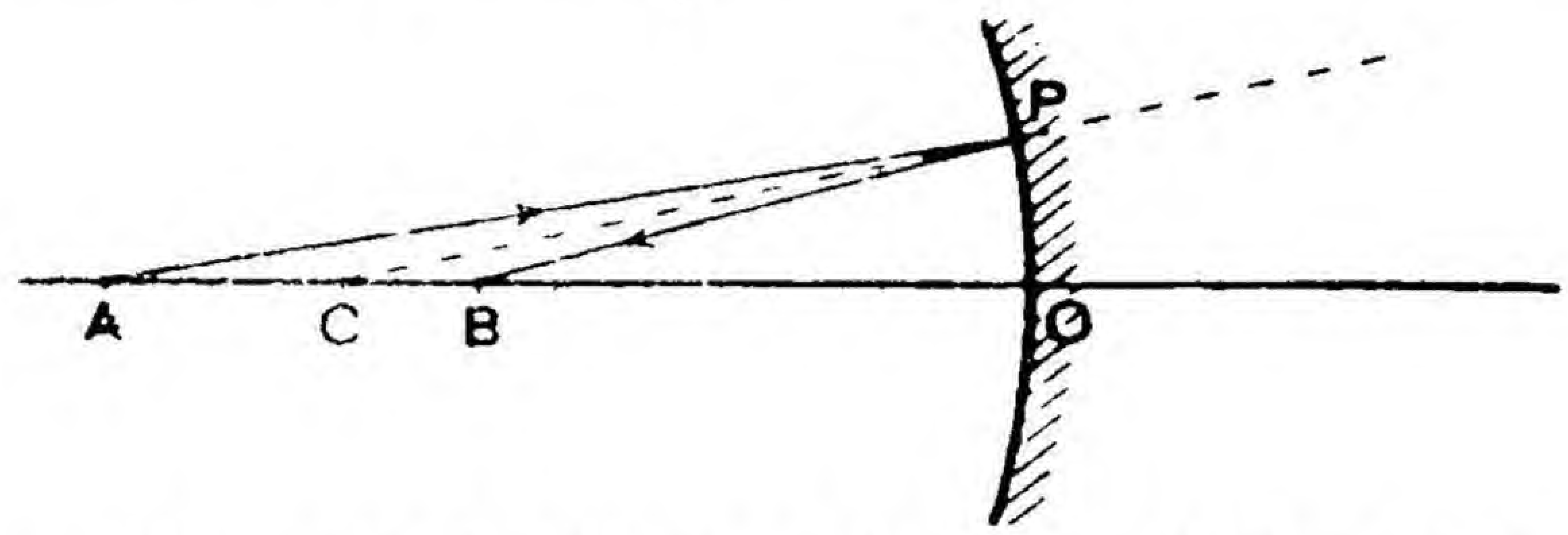
It is essential that the reader should verify the truth of this last equation in each of the cases drawn by converting it to an equation in distances and seeing that it reduces to the previous equation in every case.

$$\therefore \quad ur-uv=uv-vr$$

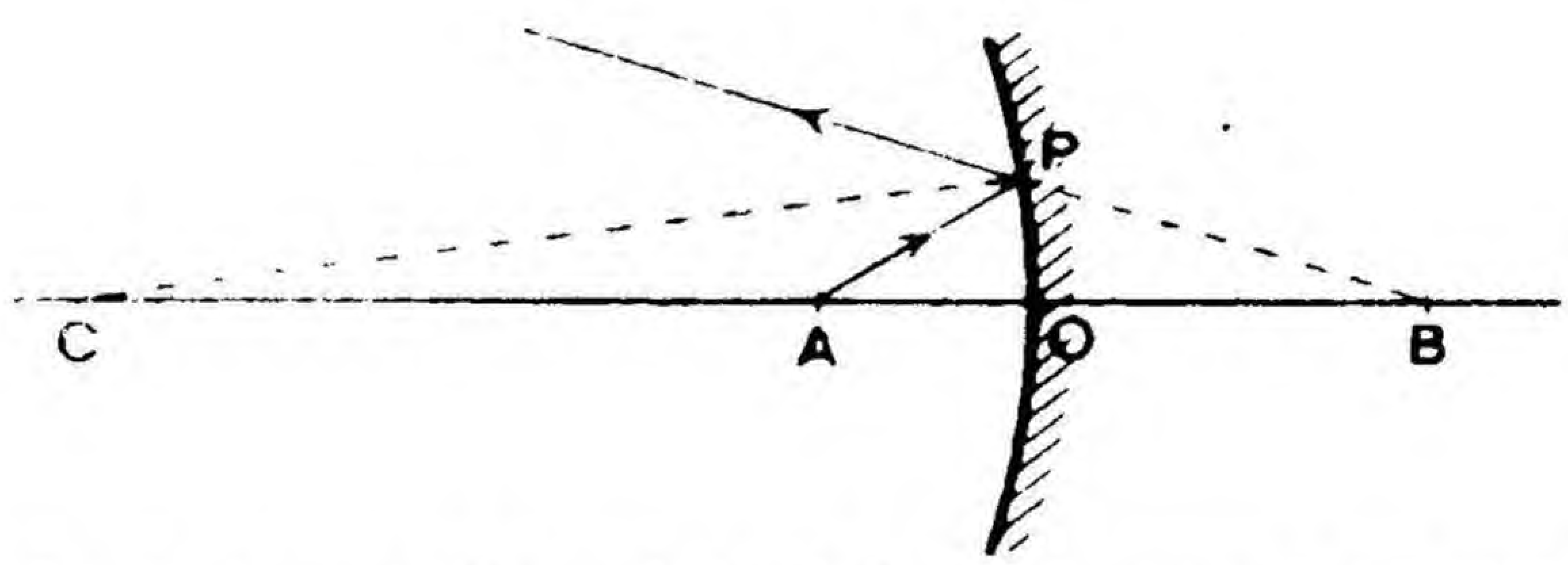
Dividing by uvr we have equation (17) again

$$\frac{1}{v}+\frac{1}{u}=\frac{2}{r}$$

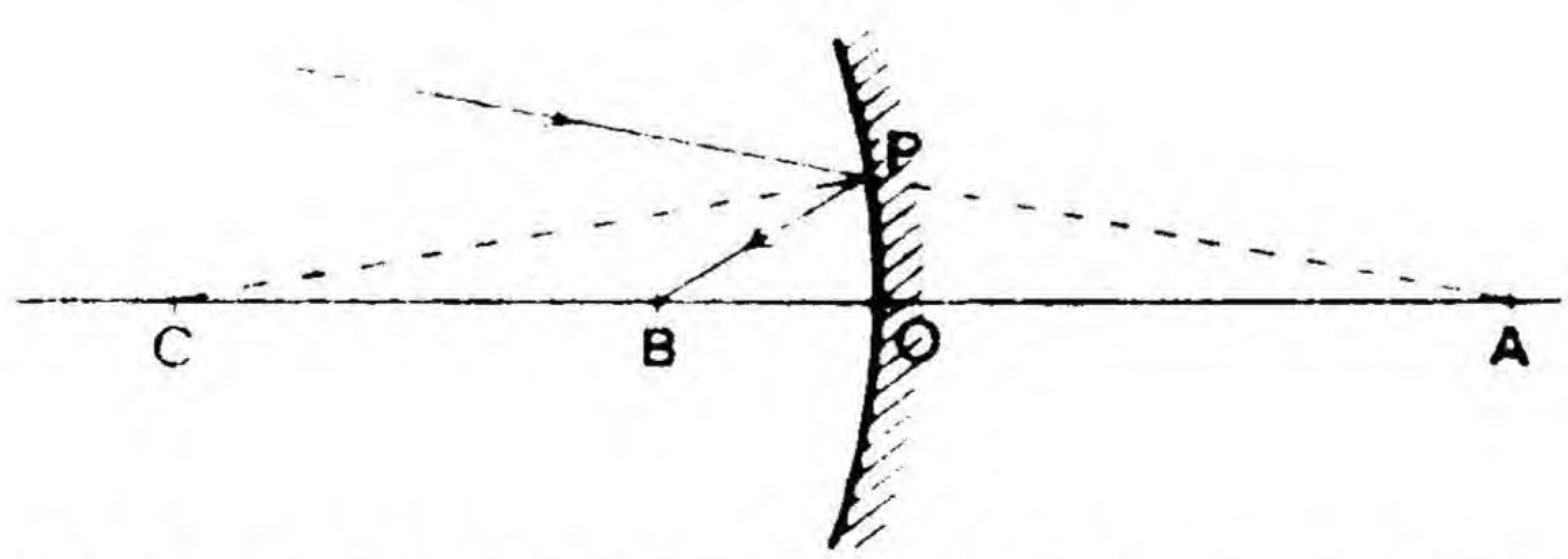
This equation shows that, given u and r, the value of v is independent of the angle which the ray from the object makes with the axis. There-



A real image of a real object formed by a concave mirror



A virtual image of a real object formed by a concave mirror



A real image of a virtual object formed by a concave mirror

Fig. 36—*continued.*

fore all paraxial rays in the plane of the diagram from the point A on the axis of a spherical mirror pass through the same point B on the axis after reflection and since conditions are symmetrical about OC as axis, paraxial rays from A in any other plane will be reflected to pass through B. Therefore a solid pencil of paraxial rays from A will be reflected to pass through B and so we have theorem 7.

Theorem 7 : a spherical mirror forms a point image of a point object on its axis.

If $u=\infty$, from equation (17)

$$v=\frac{r}{2}$$

so we have theorem 8.

Theorem 8 : rays parallel to the axis of the mirror are reflected so as to pass through a point on the axis lying half-way between the centre of curvature and the pole of the mirror. This point is called the **focus** of the mirror and its position relative to the pole is known as the focal length of the mirror and is denoted by f.

$$\therefore \quad f=\frac{r}{2}$$

and so equation (17) may be written

$$\frac{1}{v}+\frac{1}{u}=\frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

which should be compared with the similar equation for lenses. The cases of a beam parallel to the axis striking a concave and convex mirror are shown in Fig. 37, and it is obvious that the focal length of a convex mirror is positive while that of a concave mirror is negative.

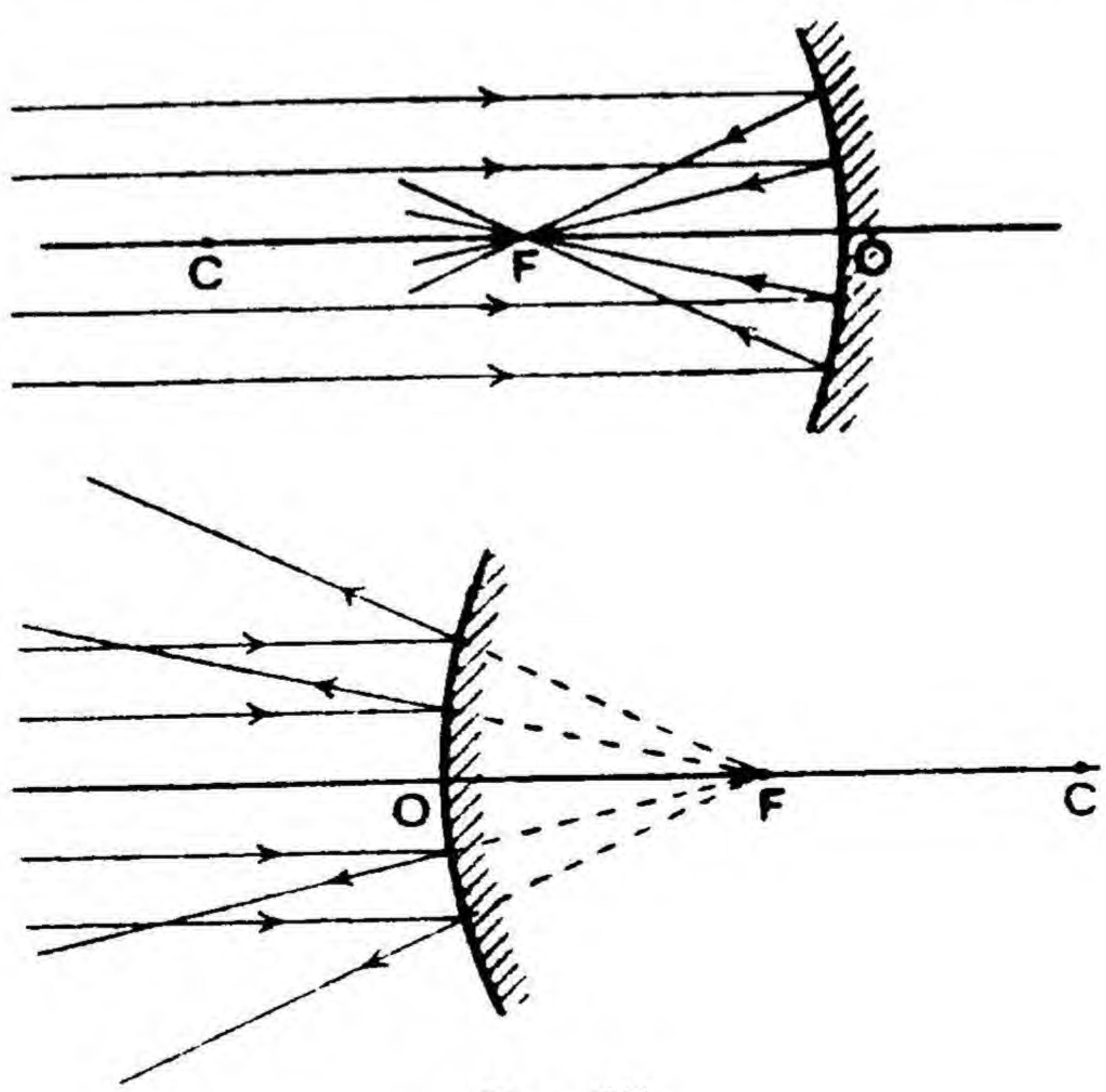


Fig. 37.

It follows from the principle of the reversibility of rays of light or by putting $v=\infty$ in equation (18), that, if the object is at the focus of the mirror, then the rays from it are parallel to the axis after reflection.

The other interesting set of rays is the one which passes through the centre of curvature of a mirror or is converging to it in the case of a convex mirror. They strike the mirror normally and so they retrace their path after reflection. So we have theorem 9.

Theorem 9 : rays passing through the centre of curvature of a mirror retrace their path after reflection. They correspond in a way to the rays going through the centre of a lens which are undeviated by the refraction. In each case, too, the corresponding point is the one place where the object and image coincide.

23. FINITE OBJECTS

The case of finite objects can be dealt with by considering the image B of a point object A on the axis of the mirror (Fig. 38). If the figure is rotated about the centre of curvature C through a small angle, A describes the small arc AA′ and B the small arc BB′. It follows at once that B′ is the image of A′, since the direct proof given for the points A and B applies equally well to A′ and B′. But the angle of rotation must be

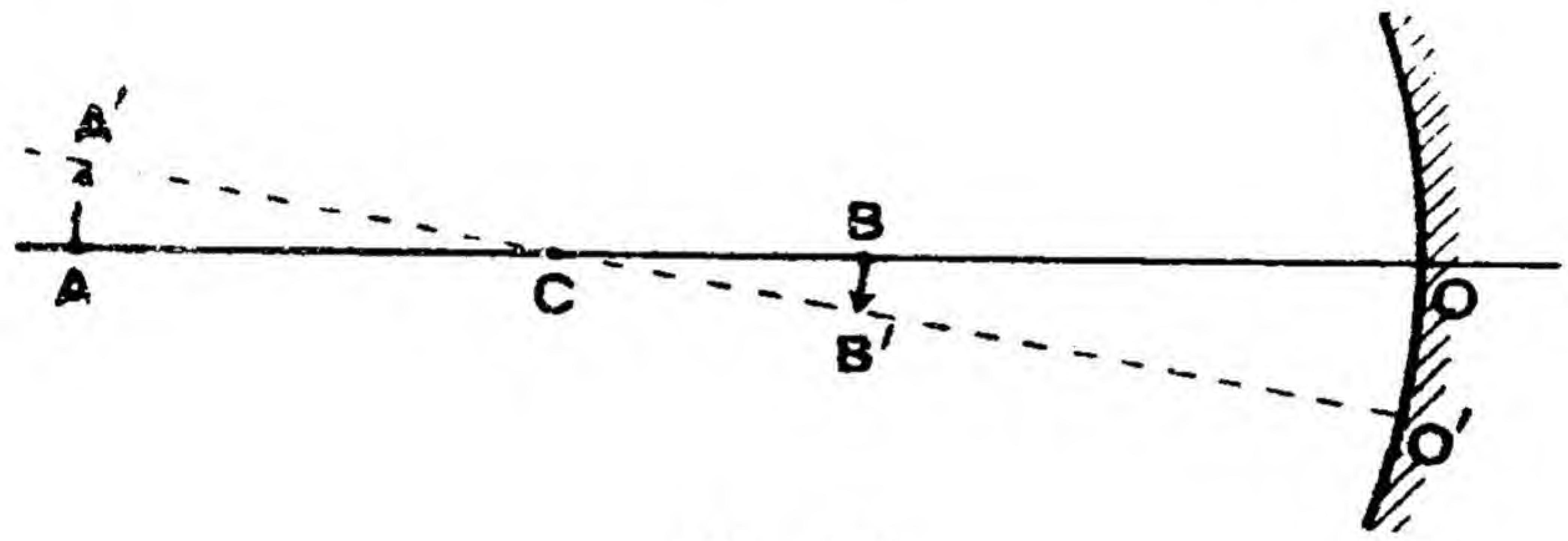


Fig. 38.

small for paraxial conditions to be satisfied and under these circumstances AA′ and BB′ are straight lines at right angles to the axis of the mirror. So we have theorem 10.

Theorem 10 : spherical mirrors form line images perpendicular to the axis of small line objects perpendicular and close to the axis ; and it follows also that they form plane images perpendicular to the axis of small plane objects perpendicular and close to the axis.

The same argument may be applied to see what will happen to a parallel beam inclined to the axis of the mirror. If Fig. 37 is rotated about the centre of curvature C through a small angle, the rays parallel to the axis become a parallel beam inclined at that small angle to the axis and the focus F describes a small arc FF′. It follows that the beam comes to a focus at F′, a point in the focal plane of the mirror, which is a plane through the focus perpendicular to its axis. The point in the focal plane at which any parallel beam comes to a focus is the point where that particular ray of the beam passing through the centre of curvature of the mirror intersects the focal plane. By the law of the reversibility of rays of light, it follows that, if a point source is placed in the focal plane of a mirror, the rays from it will be reflected as a beam parallel to the line joining the point source to the centre of curvature of the mirror.

24. THE PROPERTIES OF MIRRORS

We can now deduce the properties of mirrors either by means of the equation

$$\frac{1}{v}+\frac{1}{u}=\frac{1}{f}$$

or by means of a graphical method similar to that used for lenses. We will consider the case of a real object inside the focus of a concave mirror. Here $f=-d$, $u=-a$, $a<d$. Now

$$\frac{1}{v}+\frac{1}{u}=\frac{1}{f}$$

Substituting the above values of f and u in this equation, we have

$$\frac{1}{v}+\frac{1}{-a}=\frac{1}{-d}$$

$$\therefore \quad \frac{1}{v}=\frac{1}{a}-\frac{1}{d}$$

Since $a<d$, v must be positive and so a virtual image is formed behind the mirror at a greater distance than the object, as can easily be deduced from the last equation.

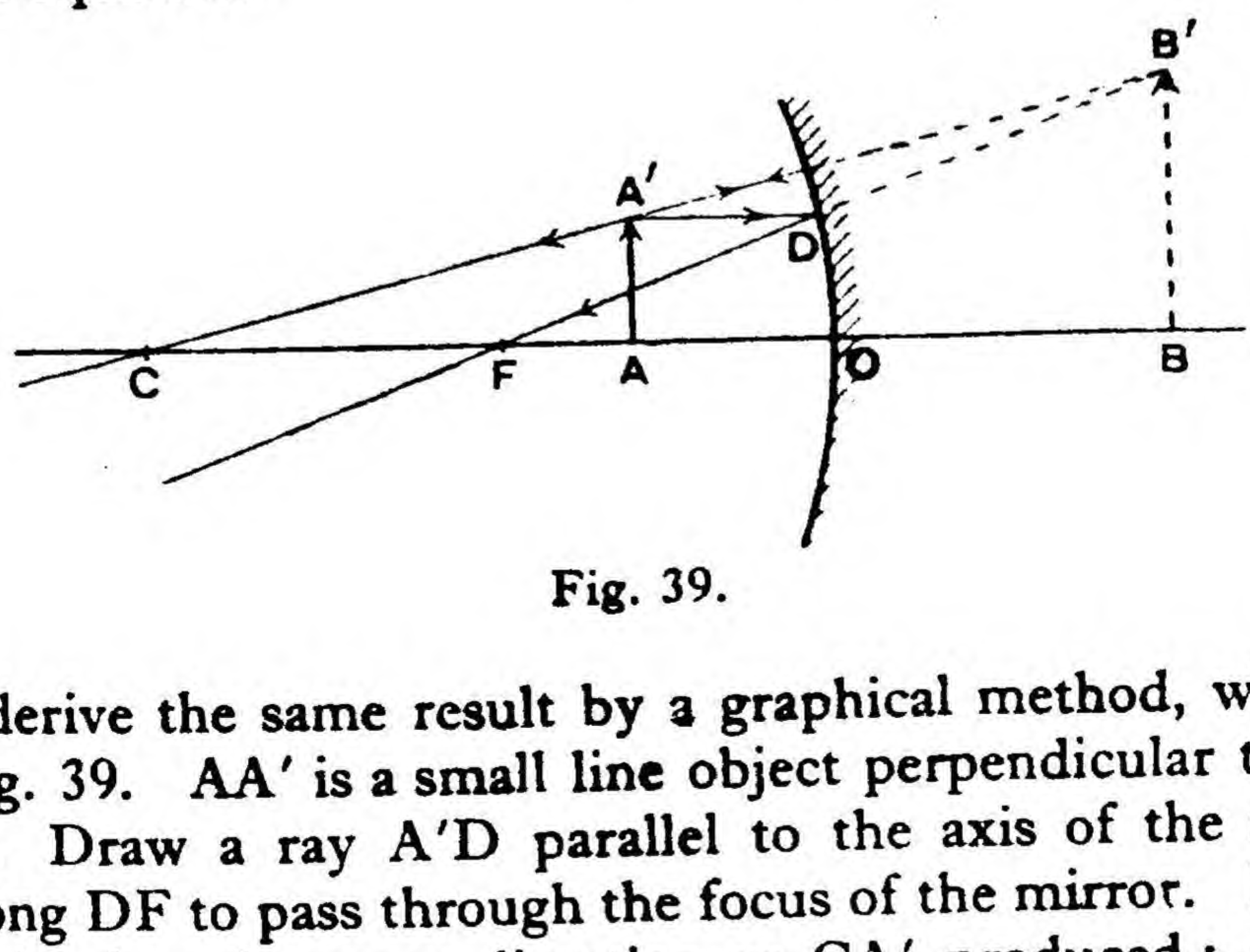


Fig. 39.

We can derive the same result by a graphical method, which is illustrated in Fig. 39. AA′ is a small line object perpendicular to the axis of the mirror. Draw a ray A′D parallel to the axis of the mirror; it is reflected along DF to pass through the focus of the mirror. Draw the ray from A′ which has the same direction as CA′ produced; it strikes the mirror normally and so retraces its path after reflection. Since the object is inside the focus, it is obvious that these two reflected rays are diverging; produce them backwards to meet at B′ behind the mirror. Since we have already proved that spherical mirrors form point images of point objects near the axis, therefore all paraxial rays from A′ necessarily appear to come from B′ after reflection, and so B′ is the image of A′. Since AA′ is perpendicular to the axis, its image must also be perpendicular to the axis, and if B′B is drawn at right angles to OC to cut it at B,

BB′ is the virtual image of AA′. The diagram shows that the image is also erect and magnified. The fundamental equation for all mirrors can be proved from this diagram by considering the two pairs of similar triangles CAA′, CBB′ and FOD, FBB′.

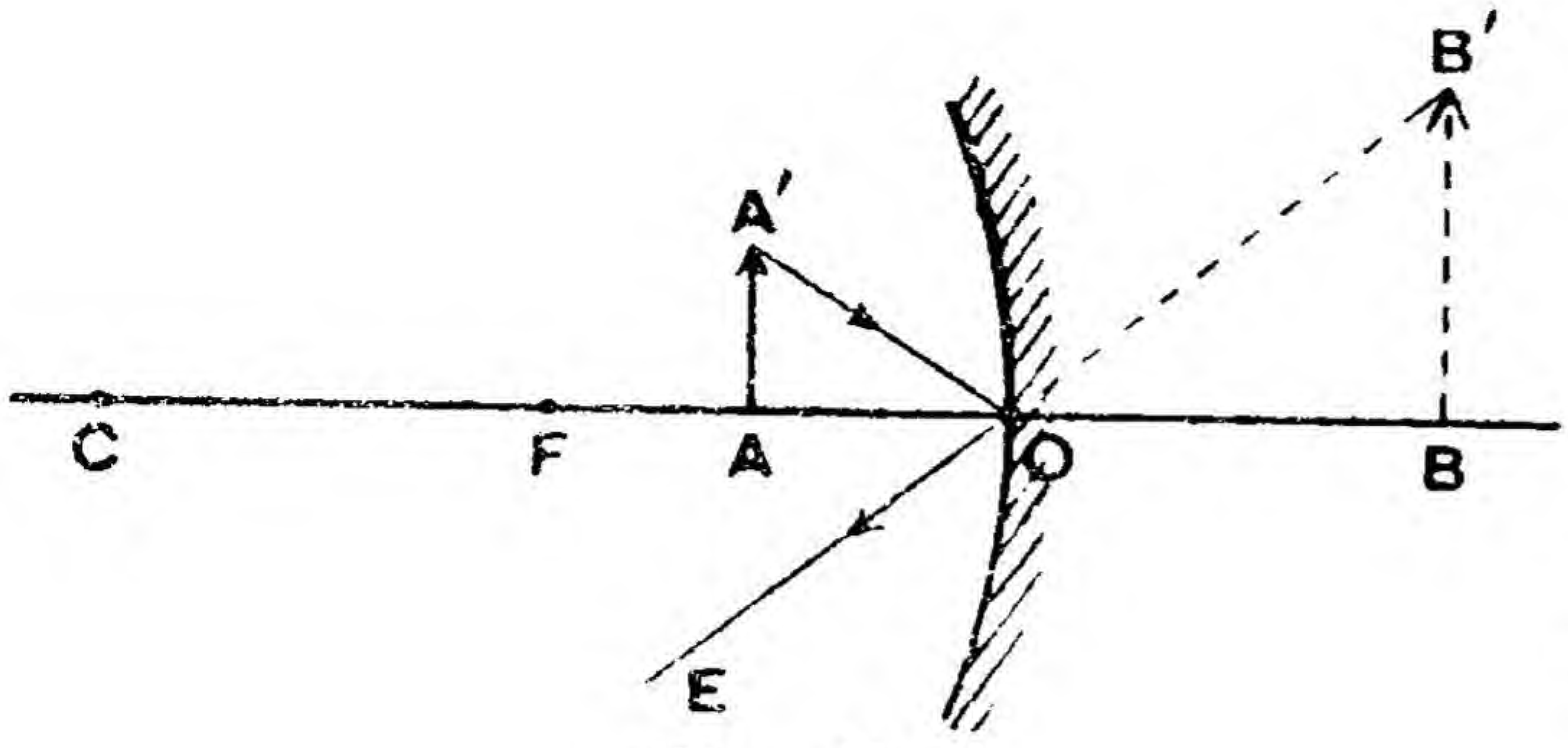


Fig. 40.

We can also obtain an expression for the magnification of a mirror from a similar diagram (Fig. 40) by considering the ray A′O which is reflected along OE so that B′OE is a straight line. It follows from this fact and the law of reflection that ∠B′OB=∠EOA and ∠EOA=∠A′OA. So ∠A′OA=∠B′OB and the triangles A′OA and B′OB are similar. Therefore the magnification, m, is given by

$$m=\frac{BB'}{AA'}=\frac{OB}{OA}=-\frac{v}{u} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

The remaining cases of interest are shown in Table 3 and the reader should verify the results both by the use of the equation and by the graphical method. In particular it is important that he should satisfy himself that the sign as well as the magnitude of the above expression for the magnification is correct for each of these cases. The case of a convex mirror forming a virtual image of a virtual object outside its focus is illustrated in Fig. 41.

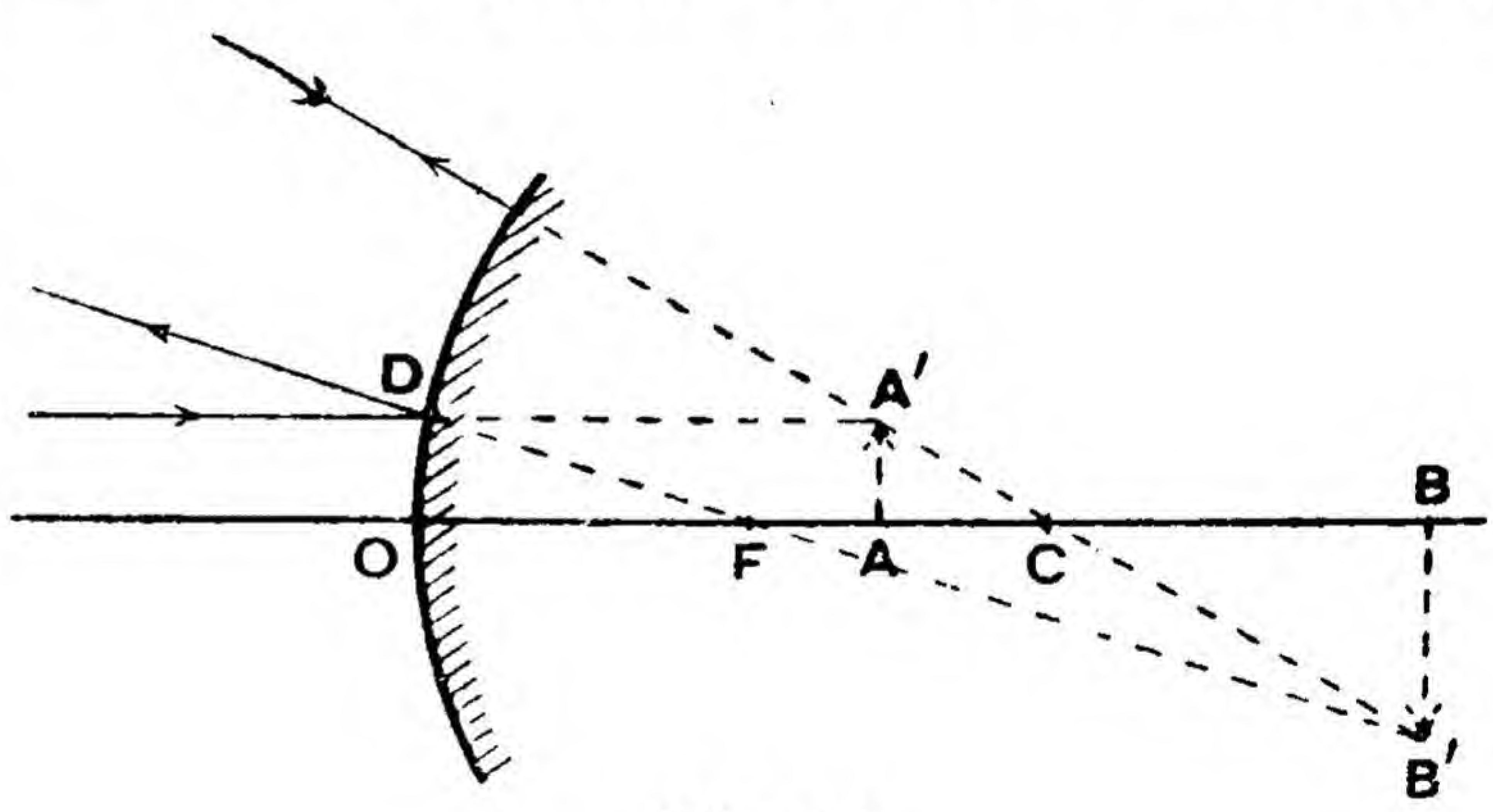


Fig. 41.

The reader should examine the table carefully and he will notice the resemblance between the effect of a concave mirror on real objects and that of a convex mirror on virtual objects ; the same similarity is noticed in the case of the effect of a concave mirror on virtual objects and that of a convex mirror on real objects. Finally, if the reader compares Tables 1 and 3, he will see that the properties of a concave mirror are the same as those of a converging lens, while those of a convex mirror are the same as those of a diverging lens.

25. EXPERIMENTAL TESTS

How do the above theorems for mirrors stand the test of experiment ? In subjecting them to this test we must remember that they are only true for paraxial rays ; with this limitation we can say that they must

be true for they are derived from an experimentally established law by deductive logic. The theorem that all rays parallel to the axis of a mirror are reflected to pass through a point on its axis can easily be verified by some kind of apparatus for producing rays of light such as a smoke-box

TABLE 3

Nature of Mirror.	Nature and Position of Object.	Nature and Position of Image.
Concave.	Real and outside C.	Real, inverted, diminished, and between C and F.
Concave.	Real and at C.	Real, inverted, the same size as the object, and in the same position.
Concave.	Real and between C and F.	Real, inverted, magnified, and outside C.
Concave.	Real and inside F.	Virtual, erect, magnified, further from the mirror than the object.
Concave.	Virtual in any position.	Real, inverted, diminished, and nearer to the mirror than the object.
Convex.	Real in any position.	Virtual, erect, diminished, and inside the focus.
Convex.	Virtual outside the focus.	Virtual, inverted, and outside the focus.
Convex.	Virtual inside the focus.	Real, erect, magnified, and further from the mirror than the object.

TABLE 4

A CONCAVE MIRROR

u cm.	v cm.	$\frac{1}{u}+\frac{1}{v}$ cm^{-1}
11·9	77·1	0·0970
13·0	50·9	0·0967
13·9	40·4	0·0968
14·9	34·4	0·0962
15·9	29·9	0·0965
16·9	26·9	0·0962
17·9	24·7	0·0962
		Mean : 0·0965 ± 0·0003

$f = 10{\cdot}38 \pm 0{\cdot}03$ cm.

or a Hartel disc. In the same way it can be shown that all rays passing through another point on the axis, the centre of curvature, retrace their path after reflection. If measurements are taken, it can be shown approximately that the focus is half-way between the centre of curvature and the

pole of the mirror. The word approximately is used advisedly here, since the rays used are by no means Euclidean lines and the positions of the focus and centre of curvature cannot be fixed to more than two or three millimetres.

It is common knowledge that mirrors form sharp images of small objects perpendicular and close to the axis, and a number of the cases noticed in Table 3 are of everyday occurrence. The real, inverted, diminished image of a real object outside its focus formed by a concave mirror is used in the reflecting telescope ; the mirror brings the distant object nearer in this way and this image is then magnified by the eyepiece. The virtual image formed by the same type of mirror when a real object is inside its focus is made use of in the concave shaving mirror, while the virtual, erect, and diminished image formed by a convex mirror of a real object in any position is made use in the reflecting mirror which must be attached to every motor car in this country. And the correctness of the properties of mirrors summarised in Table 3 can be verified in the laboratory by using an optical bench, a suitable object such as an illuminated gauze, and a screen on which to project the image.

Finally, the reader will do a number of quantitative experiments with mirrors to see to what extent they satisfy the equation $\frac{1}{v}+\frac{1}{u}=\frac{1}{f}$. A typical set of results is given in Table 4 which shows that it is true to 1 in 300.

We see then that, as in the case for lenses, our theorems concerning reflection at curved surfaces are satisfied by actual mirrors and so our analysis of both lenses and mirrors for paraxial rays will be a very useful guide in enabling us to predict new uses of these instruments. Our theorems are usually more closely obeyed by mirrors than by lenses ; this is just what we should expect, because we made the assumption in dealing with the latter that the poles of the two refracting surfaces were coincident, which is never the case for an actual lens. No such assumption was made in the case of mirrors.

26. MIRRORS OF LARGE APERTURE

We have so far confined our attention to paraxial rays almost entirely, but it is evident that there are many optical instruments in which such conditions are not satisfied, as, for example, when we wish to photograph a large object close to a camera. The rays from the edge of the object must make large angles with the axis of the lens and our treatment breaks down. A similar example is the rays from an object which is being examined in a high-power microscope ; the rays from the object to the edge of the objective make quite a large angle with the axis of the microscope. And the diameter of the mirror of the large reflecting telescope at Mount Wilson is 100 inches, and so paraxial conditions are certainly not satisfied by those rays from a star which strike the edge of the mirror.

And a mirror of twice that diameter is being prepared at the moment! The full investigation of the effects to be expected for rays making large angles with the axis will be postponed to a later chapter, but, as this problem is so important, we shall consider it non-mathematically in this chapter, confining ourselves to reflection.

There are two quite separate problems to be solved. Firstly, what happens if we restrict ourselves to small objects near to the axis of the mirror, whose diameter is so large that rays making large angles with its axis can be reflected from it? Secondly, what happens if we have small objects lying a long way from the axis of the mirror?

We will take the first problem. By restricting ourselves to paraxial conditions, we have really been confining ourselves to mirrors of small aperture. The aperture of a mirror, or a lens, is the ratio of the diameter of its circular outline to its focal length. If a camera lens is stopped down to $\frac{f}{11}$, its effective aperture is $\frac{1}{11}$. If a mirror has a diameter of 8 cm. and a focal length of 30 cm., its aperture is $\frac{4}{15}$. We are now going to consider the effect of a mirror of large aperture on an object. Let us

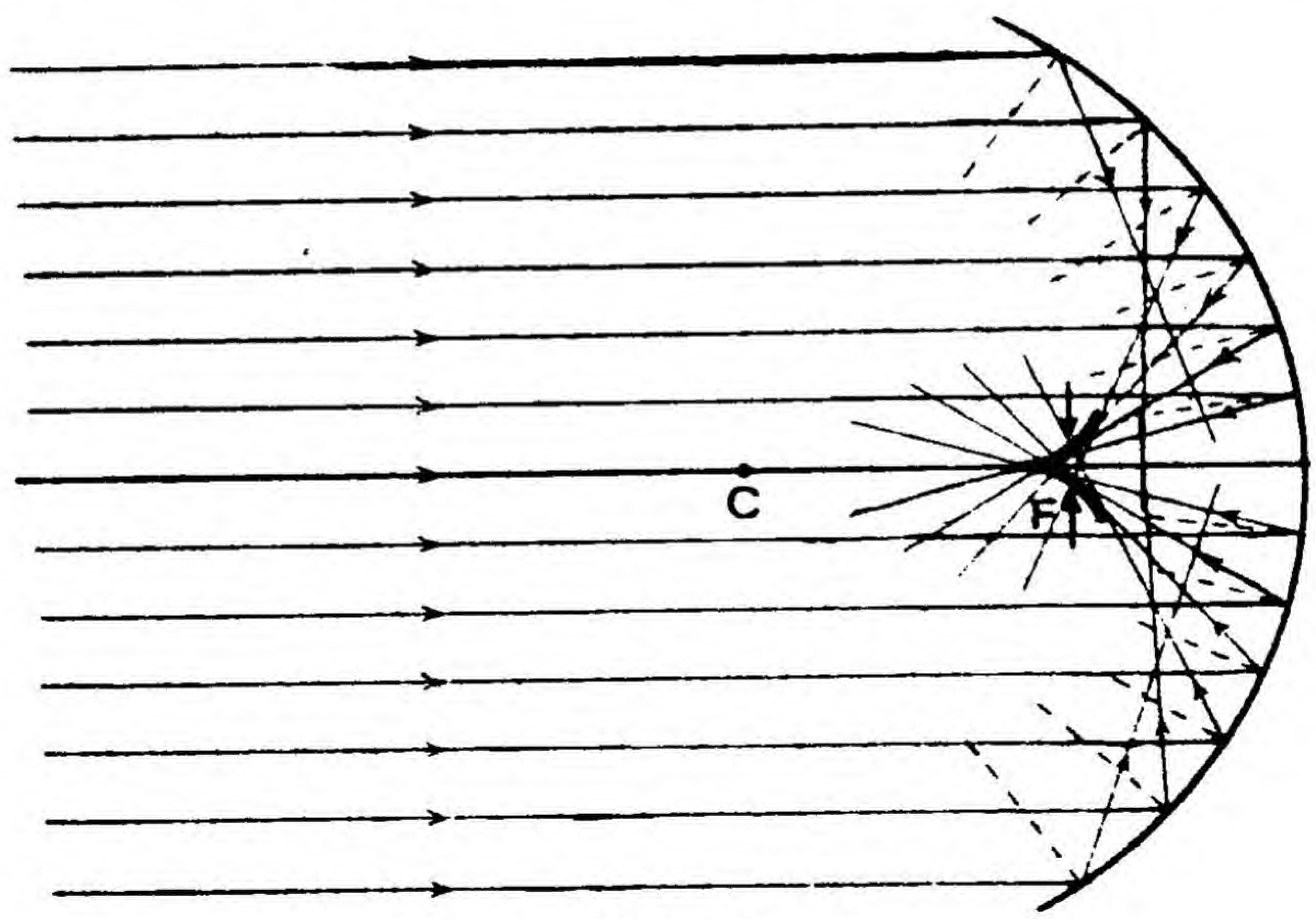


Fig. 42.

take the simplest object there is, a point object on the axis of the mirror at infinity. Is a point image formed? We know that the paraxial rays come to a focus at a point half-way between the pole of the mirror and its centre of curvature, but what about the others? We could calculate what will happen to them, but we will content ourselves for the moment by actually drawing the path of the reflected rays by assuming the law of reflection (Fig. 42). We see that the rays do not come to a point focus at all, but that there is a region of maximum concentration of the rays or of maximum illumination. It lies along a curve called a **caustic**, whose apex or cusp, as it is called, lies at the focus of the mirror. Almost everyone is familiar with this caustic, for it can be seen in a glass of water on

which the sun is shining. It is hardly true to describe the sun as a point source, but the rays from the top of the sun are only inclined to those from the bottom at an angle of half a degree, and so the collection of rays from the sun can be regarded as one parallel beam for a qualitative observation of this sort. A similar caustic is produced by the glass acting as a lens and so our theorems for lenses break down too when paraxial restrictions are removed. So mirrors and lenses of large aperture do not form point images of point objects on the axis, and therefore they will not form sharply focussed images of even small plane objects in the neighbourhood of the axis. Fig. 42 is confined to one plane and to obtain the result of the reflection of a solid pencil of rays it must be rotated about the axis of the mirror through 180°, when the caustic curve becomes a kind of cone whose vertex is at the focus of the mirror. Since all the rays at the same distance from the axis pass through the same point on the axis after reflection, there is also a bright line running from the focus F inwards along the axis. The caustic seen by reflection in a glass is the intersection of the cone and the bottom of the glass, the bright line being absent since the glass is cylindrical, not spherical. The best image which we can project on to a plane surface is a circle of finite radius called the **circle of least confusion**, which is obtained by putting the screen in the position shown by the two arrows pointing inwards. This is the sharpest image which can be obtained of the point object at infinity and indeed of a point object at any distance from the mirror. The reader will see that this lack of point focus arises from the fact that the further a ray is from the axis, the nearer to the pole of the mirror does the reflected ray cross the axis. In other words, the focal length of the mirror is less the further the ray is from the axis. If the diagram is examined more closely, it can be seen that this decrease of focal length becomes more rapid the further out we get from the axis. A more general way of expressing this point is to realise that the real factor controlling the variation in focal length is the angle of incidence rather than distance of the rays from the axis, and we see that the focal length decreases with increase of angle of incidence and that the rate of decrease increases as the angle of incidence increases. We shall see the importance of this in considering the second problem.

27. IMAGE FORMATION OF OBJECTS A LONG WAY FROM THE AXIS OF A MIRROR OF SMALL APERTURE

The second way in which paraxial conditions can be violated is by using mirrors of small aperture to form images of objects a long way from the axis. Let us consider a parallel beam of rays originating in a point object at infinity and making a large angle with the axis CO of a mirror DD′, centre C, and pole O (Fig. 43). The extreme rays AD and A′D′ from the object to the edges of the mirror make large angles of

incidence, and so our simple equation for the mirror no longer holds. But it is clear that the mirror can be regarded as a small marginal portion of a mirror of large aperture of axis CO′ parallel to the incident beam, its outline being shown in dotted lines in the diagram. We can now use the results of the previous article. It follows that, after reflection, the marginal ray A′D′ will cross the axis CO′ of the imaginary mirror closer to its pole than the parallel ray AD. So DTS is the path of AD after reflection and D′TV that of A′D′. This follows from the fact that the incidence is so large that we have reached the stage where the focal length changes so rapidly with variation in angle of incidence that *there is a difference even for a mirror of small aperture.* Now in practice the pencil of rays is a solid pencil and the mirror is in three dimensions. To obtain the true state of affairs, the figure is rotated through a small angle about CO′ as axis, when the line DD′ describes the surface of the mirror of small aperture we are considering and the incident plane pencil becomes a solid pencil, as it is in practice, and the reflected pencil has the necessary shape to enable it to pass through two lines, one through T normal to the plane of the paper, and the other SV. It is not easy to describe the shape of this reflected pencil, but the reader should try to see it for himself in a smoke-box. These two lines are called the **focal lines** of the point source and they are separated by a finite distance and are mutually perpendicular. Here again the best focus that can be obtained is the circle of least confusion which is produced on a screen placed between T and SV at the position where the reflected pencil has the minimum diameter. So a mirror of small aperture will not give a sharply focussed image of a point object remote from its axis, nor of small plane objects in a similar position.

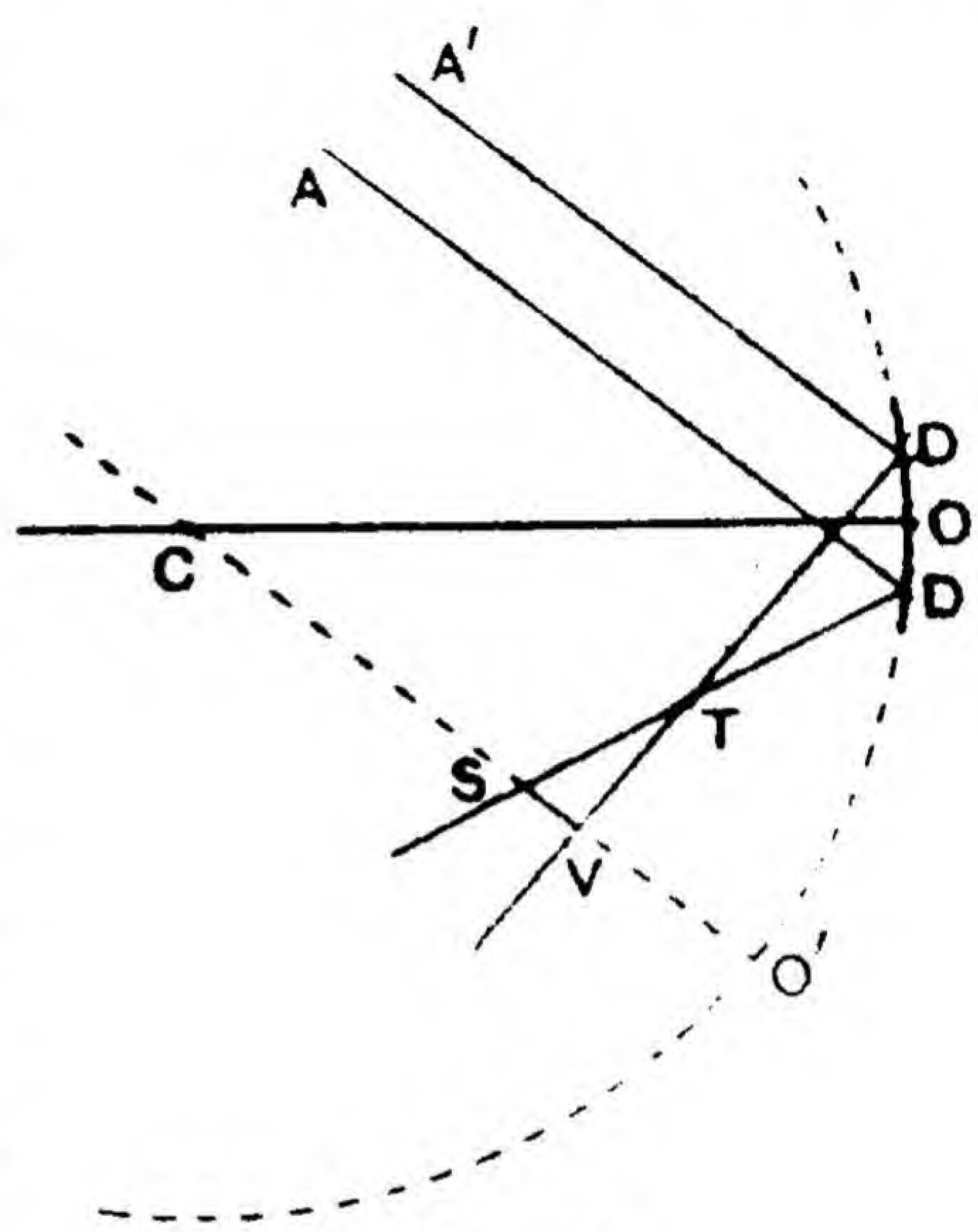


Fig. 43.

This reflected beam is another example of an astigmatic pencil, which is one without a point focus, and this result is so important that we shall consider another way of establishing it. Rays such as AD and A′D′ in the plane containing the object and the axis of the mirror not only strike the mirror obliquely, but the actual diameter of the mirror is foreshortened in a direction normal to the rays. But rays in a plane normal to that containing the object and the axis of the mirror only strike the mirror obliquely; there is no foreshortening of its diameter. The oblique incidence causes the focal length of rays confined to either plane, measured along the principal reflected ray OT, to be less than that of paraxial rays. But the foreshortening effect for rays such as AD and A′D′ causes their focal length to be less than that of rays in a perpendicular plane. There-

fore the rays are bent inwards in the plane of the diagram more rapidly than in a perpendicular plane, and the reflected pencil comes to a line focus at T normal to the plane of the diagram and to another line focus SV in the plane of the diagram. These foci are called the **tangential** and **sagittal** focal lines respectively. This way of looking at the formation of the reflected pencil connects it up with the astigmatic lens (Art. 20); we see that in each case the astigmatism arises from the fact that the optical system has different focal lengths in two mutually perpendicular planes. In the astigmatic lens this is due to the asymmetrical nature of one of its surfaces, here it is due to the asymmetrical position of the object. While the point object is on the axis of the mirror, spherical symmetry is maintained, but when it is to one side the symmetry is lost and the mirror has different focal lengths for the plane containing the point object and the axis and a perpendicular plane. So we see that the image of small plane objects remote from the axis of a spherical mirror of small aperture will suffer from the same defects as those formed by an astigmatic lens. They will never be sharply in focus and if they should be cross-wires or a gauze, then only one set of lines can be focussed at a time. In short, even if we stop down a mirror of large aperture to maintain the sharpness of focus of those parts of an object near to its axis, the sharpness of focus or definition will still get progressively worse as we go outwards from the axis.

We may sum up the present position of our analysis in this way: starting from the laws of reflection and refraction of rays of light, we have been able to deduce a number of theorems concerning thin lenses and mirrors. These theorems have only been proved for paraxial conditions, and under these conditions it is found experimentally that actual lenses and mirrors obey them and so they form a sound basis for a prediction of the properties of actual lenses and mirrors. But a preliminary enquiry into what happens when paraxial restrictions are removed has shown us that the above theorems are no longer valid, and if we wish to use lenses of large aperture so as to make our images brighter or to deal with large objects we shall have to extend our investigations. Before doing this we shall enquire into another defect of lenses which puzzled Newton and led to some epoch-making investigations into the nature of white light. Anyone who has a simple magnifying glass can see that, if he pushes it to the limit of its magnification, the images get blurred and slightly *coloured* at the edges. Newton had noticed the same thing in the case of the telescopic objectives of his time, and he decided to investigate the nature of white light itself to see if that was the cause of the trouble. We cannot do better than follow Newton in his investigation, postponing our consideration of the other problems until later.

EXAMPLES ON CHAPTER III

1. Show that if a beam of rays diverging from a point near the axis strikes a small concave mirror, the rays after reflection will pass through another point. How is the position of the second point related to that of the first ? If the first point is 17 cm. from the surface of the mirror and the second is 17 cm. from the centre of the curvature, what is the radius of curvature of the mirror ? (*London Inter.*)

2. Why is a sign convention used in geometrical optics ? Illustrate your answer by deducing, from first principles, the equation connecting the distances u and v of a real object and its image respectively from a concave mirror of radius r, (a) when the image is real, (b) when the image is virtual, and then applying the convention you yourself employ.

A convex mirror, used in a car as a rear view reflector, reflects into the eye placed centrally and 20 inches away rays incident on the periphery of the mirror and inclined at 10° to the central axis. If the mirror is 3·5 inches across determine approximately its radius of curvature. (*N.U.J.B.*)

3. A man requires a mirror for shaving to produce an image of his face twice the actual length of his face when he is 10 in. from the mirror. What sort of mirror must he use, what will be its focal length, and how far away from the mirror will the image of his face be formed ?

4. A dentist uses a concave mirror for examining his patients' teeth. Design a mirror suitable for this purpose, assuming that the image is to be three times the size of the object and the mirror is not to be held nearer to the tooth than 0·5 in.

5. A mirror hangs on the wall of a room and produces a virtual erect image 6 in. long of a man 6 ft. high standing against the opposite wall 24 ft. away. Find the position of the image and the radius of curvature and nature of the mirror.

6. Why is the driving mirror of a motor car convex rather than concave ?

Such a mirror produces an image 1 in. wide of a car 3 ft. wide when it is 30 yds. away from the mirror. Find the radius of curvature of the mirror. How can the driver tell, by looking into the mirror, whether the car is overtaking him or receding from him ?

7. A concave mirror is filled with a liquid. What measurements would you make in order to determine the refractive index of the liquid ? (*Oxford Schol.*)

8. How would you determine the position of the centre of curvature of a concave mirror using only one pin ?

An object is placed at the centre of curvature of a concave mirror, of focal length 10·25 cm., which is held horizontally. Transparent liquid of refractive index 1·6 is poured into the mirror to a depth of 0·5 cm. at its centre. In order that the position of the object and the image formed by the liquid and mirror may coincide the object needs to be moved. Explain why the movement is necessary and calculate its magnitude. (*N.U.J.B.*)

9. PBCA is the axis of a concave spherical mirror, A being a point object, B its image, C the centre of curvature of the mirror, and P the pole. Find a relation between PA, PB, and PC, supposing the aperture of the mirror to be small.

A concave mirror forms, on a screen, a real image of twice the linear dimensions of the object. Object and screen are then moved until the image is three times the size of the object. If the shift of the screen is 25 cm. determine the shift of the object and the focal length of the mirror. (*N.U.J.B.*)

10. You are given a mirror, the diameter of whose outline is 4 ft., and you cannot decide by feeling whether it is plane, convex, or concave. Describe in detail how you would decide which of the above alternatives is true.

11. A pin is placed 30 cm. in front of a converging lens of focal length 20 cm. When a convex mirror is placed 20·0 cm. behind the lens, a real inverted image of the pin is produced coinciding with the pin itself. Draw a diagram (a) for a point on the axis, (b) for a point off the axis of the lens and mirror showing exactly how this is done and calculate the radius of curvature of the convex mirror.

12. A pin is placed 30 cm. in front of a concave lens and, when a concave mirror of 40 cm. radius of curvature is placed 25·0 cm. behind the lens, a real inverted image of the pin coinciding with the pin itself is formed. Draw a diagram showing the rays from a point (a) on the axis, (b) off the axis of the lens and mirror illustrating how the image is formed and calculate the focal length of the concave mirror.

When a pin is placed 28·5 cm. in front of the lens above, a *faint* real inverted image is formed coinciding with the pin itself. What is the radius of curvature of the face of the lens nearer to the pin? When the lens is turned round a similar image is formed with the pin 31·8 cm. from the lens. What is the radius of curvature of the other face of the lens and the refractive index of its material?

13. What is meant by a caustic curve (or caustic surface) in geometrical optics? Discuss from this point of view the image of a point object formed by refraction at a plane surface separating a dense from a less dense medium, as well as the image formed when parallel light is incident axially (a) on a spherical, (b) on a parabolical concave mirror. (*Camb. Schol.*)

14. Deduce a formula connecting the distances of object and image from a spherical mirror.

What are the advantages of a concave mirror over a lens for use in an astronomical telescope?

A driving-mirror consists of a cylindrical mirror of radius 10 cm. and length over the curved surface of 10 cm. If the eye of the driver be assumed to be at a great distance from the mirror, find the angle of view. (*O. and C.*)

15. A point source of light, O, lies 15 cm. away from the pole, P, of a small concave mirror of focal length 5 cm. The direction PO makes an angle of 30° with the axis of the mirror. Show that two focal lines are formed by the reflection at the mirror of light from O, and find their positions. (*London B.Sc.*)

16. Give an account of the defects of the image produced (a) by reflection and (b) by refraction at a spherical surface. How would you demonstrate the position of focal lines in the case of reflection by a spherical concave mirror? (*London B.Sc.*)

17. What do you mean by the term refractive index?

The convex side of a plano-convex lens has a radius of curvature of 20 in., but when this side is silvered the lens behaves like a concave mirror of focal length 8 in. What is the refractive index of the lens? (*Camb. Schol.*)

18. A ray of light enters a thin glass lens in a direction parallel to the axis, and after internal reflection at the second surface is refracted out again at the first surface. Show that if the lens is a meniscus and the radii of the first and second surfaces are r and $3r$ respectively, the emerging ray will be parallel to the axis. (Refractive index of the glass $= 1·5$.) (*Camb. Schol.*)

Chapter IV

THE NATURE OF WHITE LIGHT AND THE CAUSE OF BLURRED IMAGES PRODUCED BY TELESCOPES

28. INTRODUCTORY

If the reader is to appreciate the amazing originality and accuracy of Newton's work on dispersion and colour, he must have some acquaintance with the notions of colour which were common at that time. It is interesting to try to put oneself in Newton's place and to try to see just what facts were known and still more to see which would strike his contemporaries as the important and significant ones. The most striking cases of colour were exhibited by natural objects such as green grass and other colours shown by any landscape and also the colours of natural or artificially prepared substances such as pigments. There were also the colours produced by the rainbow, which were usually faint and in any case only temporary, and the colours which were observed at the edges of a beam of light which had passed through a prism or a lens. All these colours are usually fainter than an ordinary beam of white light. Lastly, it was known that any colour could be made by a suitable mixture of red, yellow, and blue pigments. These facts were the basis of the various theories of colour ; the most prevalent notion was that colour was a property of the body itself and was something which the body gave to the white light, as it was reflected off it or passed through it. Colour, as it were, was a chemical compound of white light and something received from the body. This chemical theory was further supported by the fact that three different colours could produce any known colour if suitably mixed ; these three colours were the elementary colours like chemical elements, and the colours obtained by mixing them were akin to chemical compounds. Again, the fact that coloured light was usually fainter than white and occurred at the edge of a shadow when it was produced by prisms and lenses led to the view that colour was a mixture of light and darkness, the precise colour depending on the proportion of light and darkness in its composition. Although Descartes had given a satisfactory quantitative explanation of the rainbow, attributing the primary bow to one reflection and two refractions and the secondary bow to two reflections and two refractions in spherical raindrops, he was

quite unable to explain the appearance of the colours and why they were in one order in the primary bow and the opposite order in the secondary bow. It was left to Newton to realise that this phenomenon really contained the true explanation of the nature of colour and he was the first person to give a complete and correct explanation of it. But he was led to interest himself in the problem of colour at the early age of twenty-four by his desire to improve the sharpness of the images formed by the object glasses of refracting telescopes. The reader must realise that the telescope was one of the great discoveries of the age, much as wireless and television are of the present age, and everyone was keenly interested in this new tool. Newton was disappointed in the lack of sharpness of the images of telescopes of large magnifying power and he noticed that the images were also coloured at the edges. As he probably realised that a lens is equivalent to a set of prisms, he decided to try the effect of a prism on white light and also on various colours. We shall now discuss Newton's experiments and the conclusions he drew from them, without necessarily describing them exactly as he carried them out, where they can be improved upon, nor retaining the order in which he himself arranged them.

29. NEWTON'S EXPERIMENTS ON DISPERSION. DIFFERENT COLOURS HAVE DIFFERENT REFRACTIVE INDICES

A horizontal strip of white paper was taken and the left-hand half was painted blue and the other half red. The paper was then viewed through a prism with its refracting edge horizontal and pointing upwards, when it was seen that the left-hand half is above the other. This proves that the blue light is deviated more than the red and so has the greater refractive index. Some black thread was then wound round the same strip of paper and a convex lens of about 3 ft. focal length was placed some 6 ft. from the paper and a screen was placed so as to receive a sharp image of that part of the thread wound round the red part of the paper. It was then found that the screen had to be moved $1\frac{1}{2}$ in. nearer to the lens to get a sharp image of the parts of the thread illuminated with blue light. This is a further proof of the fact that blue light has a greater refractive index than red light. We see how Newton did his first experiments with the commonest kind of colours and how he tried right at the outset of his investigation to connect up a physiological sensation, colour, which cannot be measured in numbers, with a physical property of light, refractive index, which is quantitative.

30. WHITE LIGHT CONSISTS OF RAYS OF DIFFERENT REFRACTIVE INDICES

Light from the sun (the only source available to Newton, who did his experiments in his rooms at Cambridge) or an arc lamp was passed

through a small hole about $\frac{1}{3}$ in. in diameter in a screen placed in front of a prism ABC whose angle is about 60° and whose refracting edge was horizontal and pointing downwards (Fig. 44). The light was deviated upwards by the prism and a pinhole image of the source of light was thrown on to a screen some 18 ft. from the prism. This image should be some $2\frac{1}{2}$ in. in diameter in the case of the sun, which subtends an angle of $\frac{1}{2}$° at the earth. To Newton's surprise, no such simple image was obtained. Instead of a circle the image was a vertical strip of light $2\frac{1}{2}$ in. wide and 10 in. long and was also coloured along its whole length, the colours being red, orange, yellow, green, blue, indigo, and violet in order of increasing deviation. This strip of light was called a **spectrum.** The fact that the image was $2\frac{1}{2}$ in. wide showed that there was no spreading out of the light in a horizontal direction, but its length of 10 in. proved that there was a spreading out of the rays in the direction in which they were deviated by the prism. Newton attributed this to the fact that the **white light consists of rays of different refractive indices** and so the prism should produce a whole set of pinhole images of the source of light, each one displaced a little relative to its neighbour according to the different refractive index of the light producing it. But this interpretation was not accepted universally. Grimaldi had suggested that the blurred images produced by lenses were due to the rays of light being diffused when they passed through the glass, each single ray being changed into a diverging pencil of rays of small angle, and it was suggested that the spreading of the rays in Newton's experiment was an example of this diffusion. Newton soon disposed of this hypothesis by placing a second prism DEF after the first one with its refracting edge vertical. If Grimaldi's hypothesis were true, the spectrum RV should have been diffused into a square patch of light, provided that the two prisms were chosen so as to produce equal spread of the light. Instead of this the spectrum was just shifted from R_1V_1 to R_2V_2 (Fig. 45), the ray which had been least refracted at the first prism being least refracted at the second one too, and similarly for the most refracted rays.

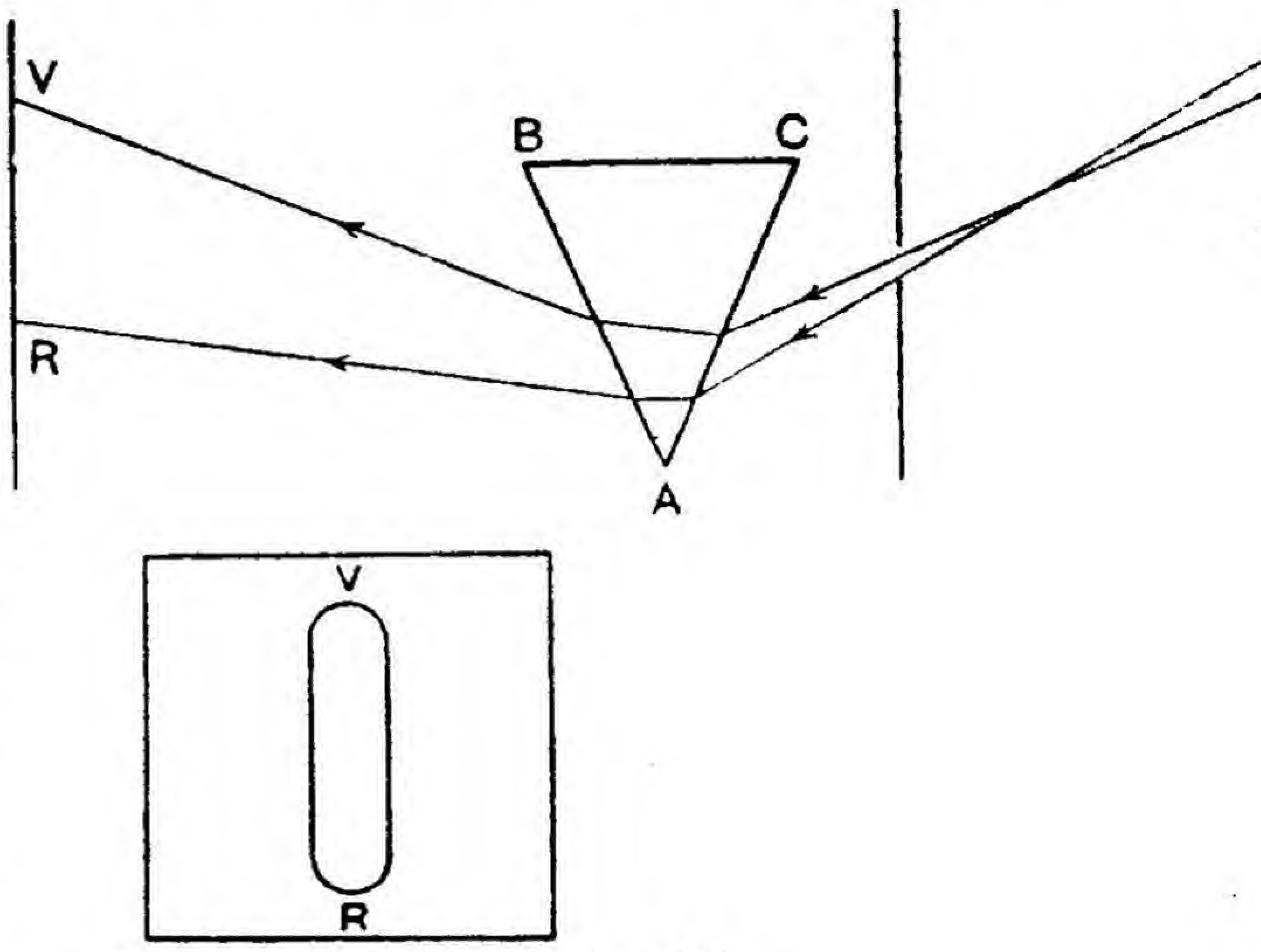


Fig. 44.

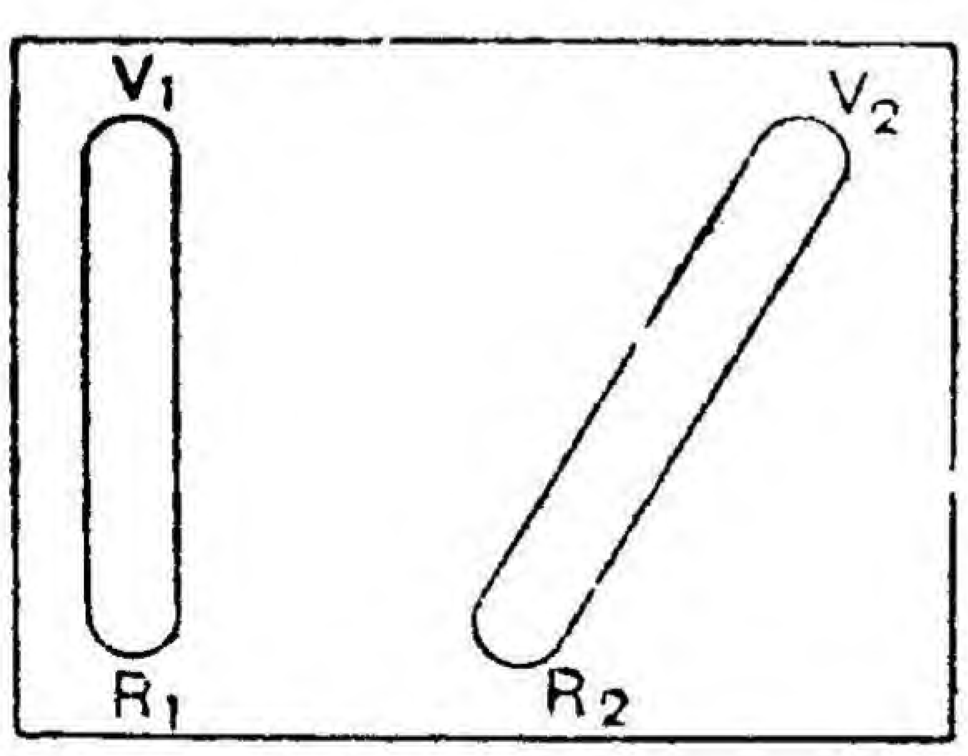


Fig. 45.

Grimaldi's hypothesis must have met with considerable support, for

Newton went to a good deal of trouble to refute it by devising variations of the above experiment. For example, he produced two spectra R_1V_1 and R_2V_2 one directly above the other (Fig. 46) and viewed them in a prism with its refracting edge vertical with the result that each spectrum was tilted as had happened in the case of the single spectrum. Again he superposed two spectra so that the red end of one coincided with the violet end of the other and vice versa, and once more viewed them in a prism with its refracting edge parallel to their length and so produced a cross (Fig. 47) as was to have been expected. These experiments were strong evidence in support of his view that the white light is composed of rays of refractive indices varying between two definite limits, since he showed that a ray of given refractive index retained that refractive index through-

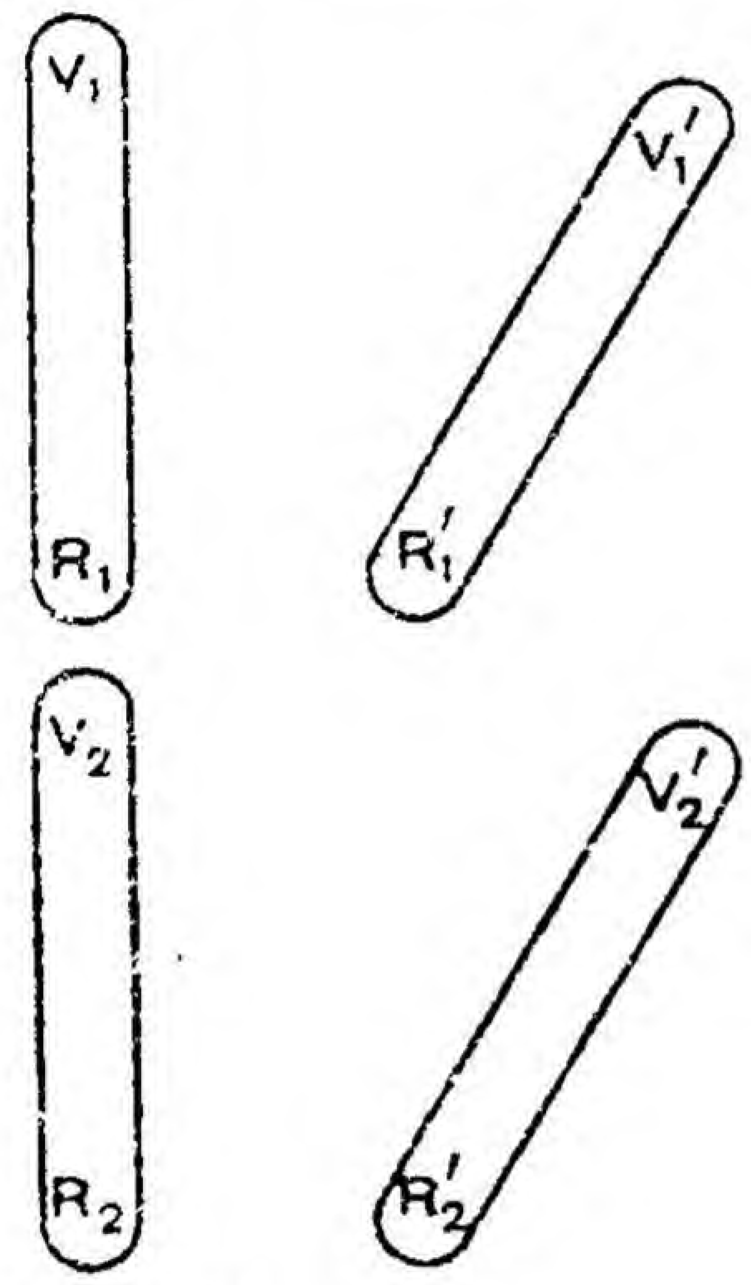


Fig. 46.

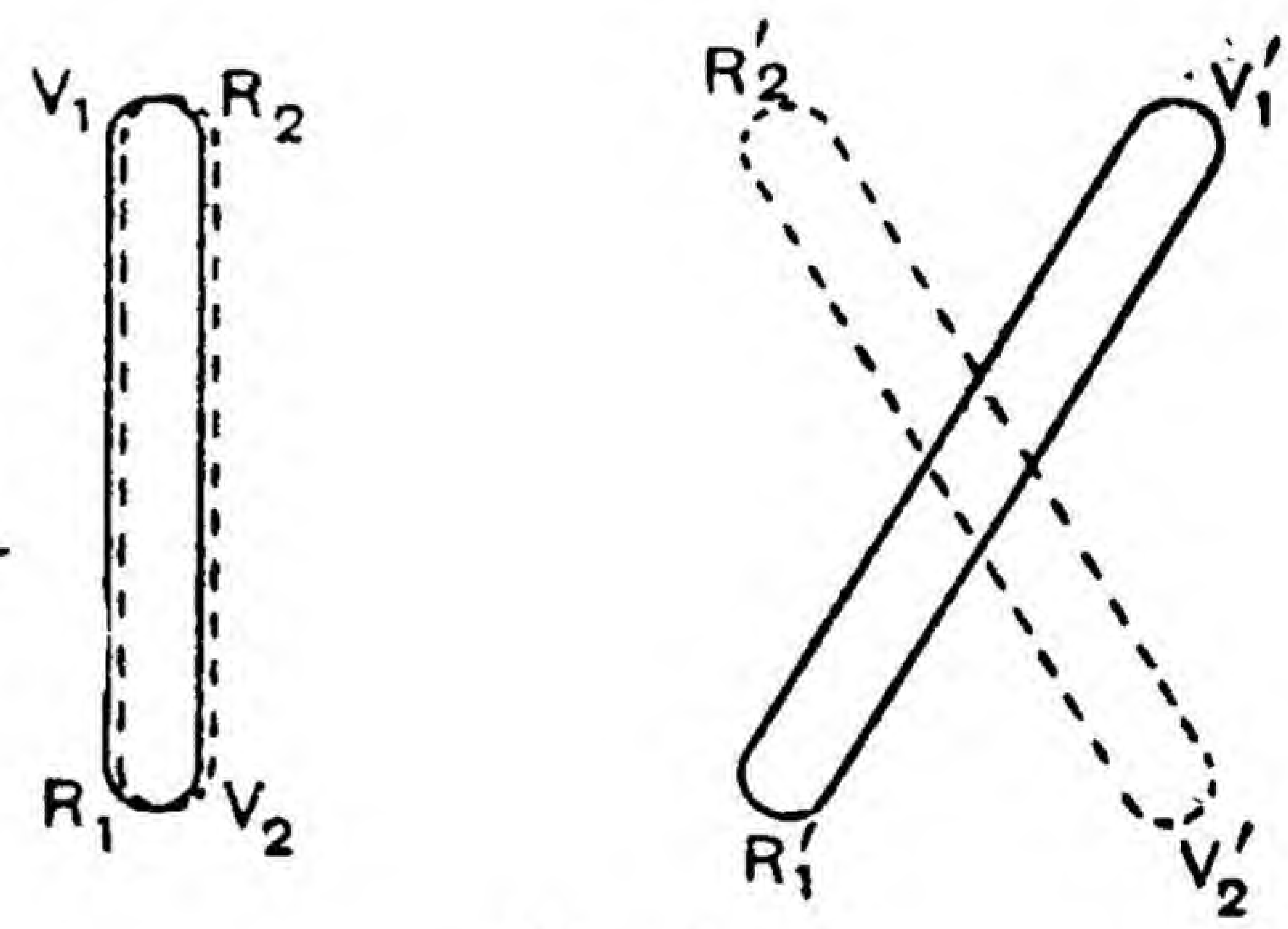


Fig. 47.

out. As yet another piece of evidence in favour of his view, he produced a spectrum as shown in Fig. 48 and arranged a board in front of a second prism so that only one colour could pass on to that prism. He kept the two boards and the second prism fixed, so that the angle of incidence should always be the same no matter what colour fell on the prism. He then rotated the first prism so as to send different colours in turn on to the second, and he was able to show that the rays which were deviated most by the first prism were so deviated by the second. On Grimaldi's hypothesis, any ray, whatever its colour, should be diffused into a divergent pencil when it passes through a prism.

His next experiment was a most significant one. It was a repetition of his very first experiment, replacing colours produced by painting a white strip of paper with spectral colours. It was clearly his intention to show that such colours are identical. So he produced two spectra and projected the violet end of one on to the top half of a vertical white strip of paper and the red end of the other on to the bottom half. He then viewed the paper through a prism with its refracting edge vertical. The upper half of the paper was deviated more than the lower half, showing that these spectral colours have different refractive indices just like natural colours

and that they retain them after reflection from paper. He got a better result by replacing the strip of paper by white thread.

As Newton believed that the blurred images produced by telescopes were due to white light consisting of rays of varying refractive index, it was natural that he should repeat his original experiment on focussing black threads wound round paper painted red and blue. This time he illuminated the printed page of a book with red light from a spectrum and, using the same lens as before, he cast a sharply focussed image of the print on a screen 6 ft. 2 in. from the lens. When he changed the illumination to violet, he had to move the screen $2\frac{1}{2}$ in. towards the lens to get the print in focus again. He remarked that this shows the difference in refractive index between the red and violet light and also that it is more marked than was the case with the painted paper. He attributed this to the fact that the colours given off by the painted paper were not pure, that is, the blue had some green and possibly yellow mixed with it, and he suggested that the distance obtained this time might also be increased if it were possible to use the extreme ends of the spectrum. This was not

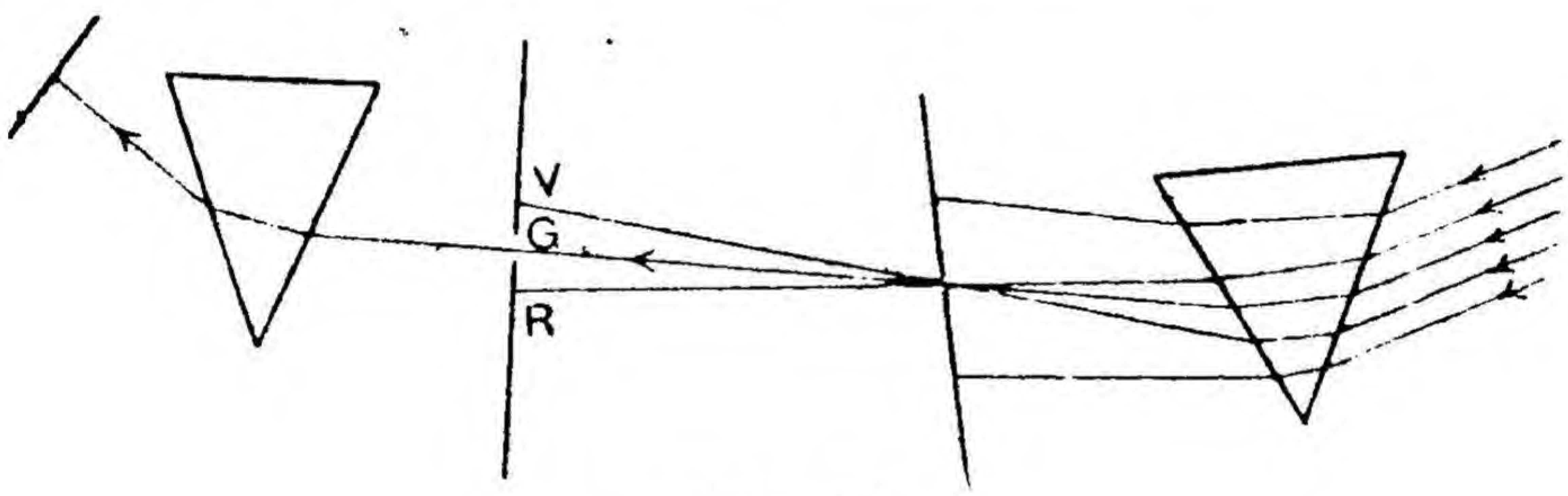


Fig. 48.

possible in the experiment, since the amount of light obtainable from the ends of the spectrum was not great enough to light the print up well enough to get an accurate focus.

So far the rays of different refractive index had been produced from colours of natural bodies or by refraction. There was no satisfactory theory of refraction at this time, and for that reason Newton felt that his interpretation of the results obtained from refraction might be criticised. This illustrates a state of affairs in scientific investigation which is very common and must be appreciated by the reader. It is quite true to say that the fundamental basis of all science is experimental fact or observation. But it is equally true that science is emphatically not a mere collection of such facts ; the science has hardly begun until the facts have been fitted into some rational scheme or working thought model, that is, until a theory has been put forward to explain the facts. The merit of the theory is that it throws new light on the facts by showing which are fundamental and which are mere matters of detail ; it lights up the whole set of phenomena, as it were, by arranging them rationally and it also suggests new and fruitful lines of investigation. Consequently Newton felt that it would support his case, if he could produce rays of

different refractive indices from white light by means of reflection, of which quite a satisfactory theory had been put forward. The reader will recollect that the law of reflection follows from the principle that a light ray passing between two points takes the shortest path. Such a principle makes an intellectual appeal and does form a confirmation of the facts about reflection. Until facts have been fitted into some kind of rational scheme, there is always a possibility that purely experimental law may be a coincidence which will break down in the face of further investigation, as was the case with Bode's law applying to the distances of the planets from the sun. So Newton performed the following two experiments. He took an isosceles right-angled prism ABC and sent a beam of white light normally on to one of the faces AC containing the right angle A (Fig. 49). Some of the light was reflected from the face BC

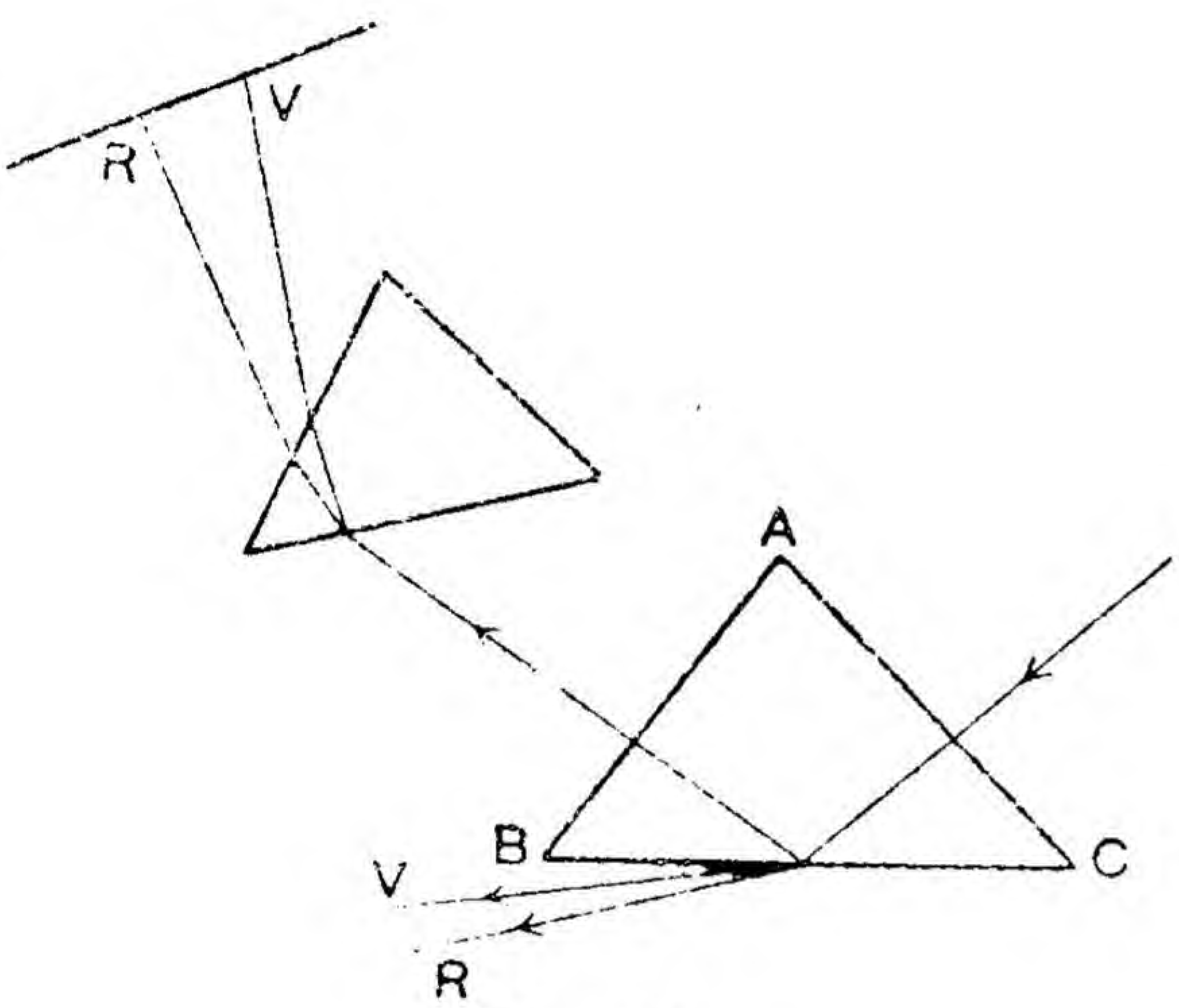


Fig. 49.

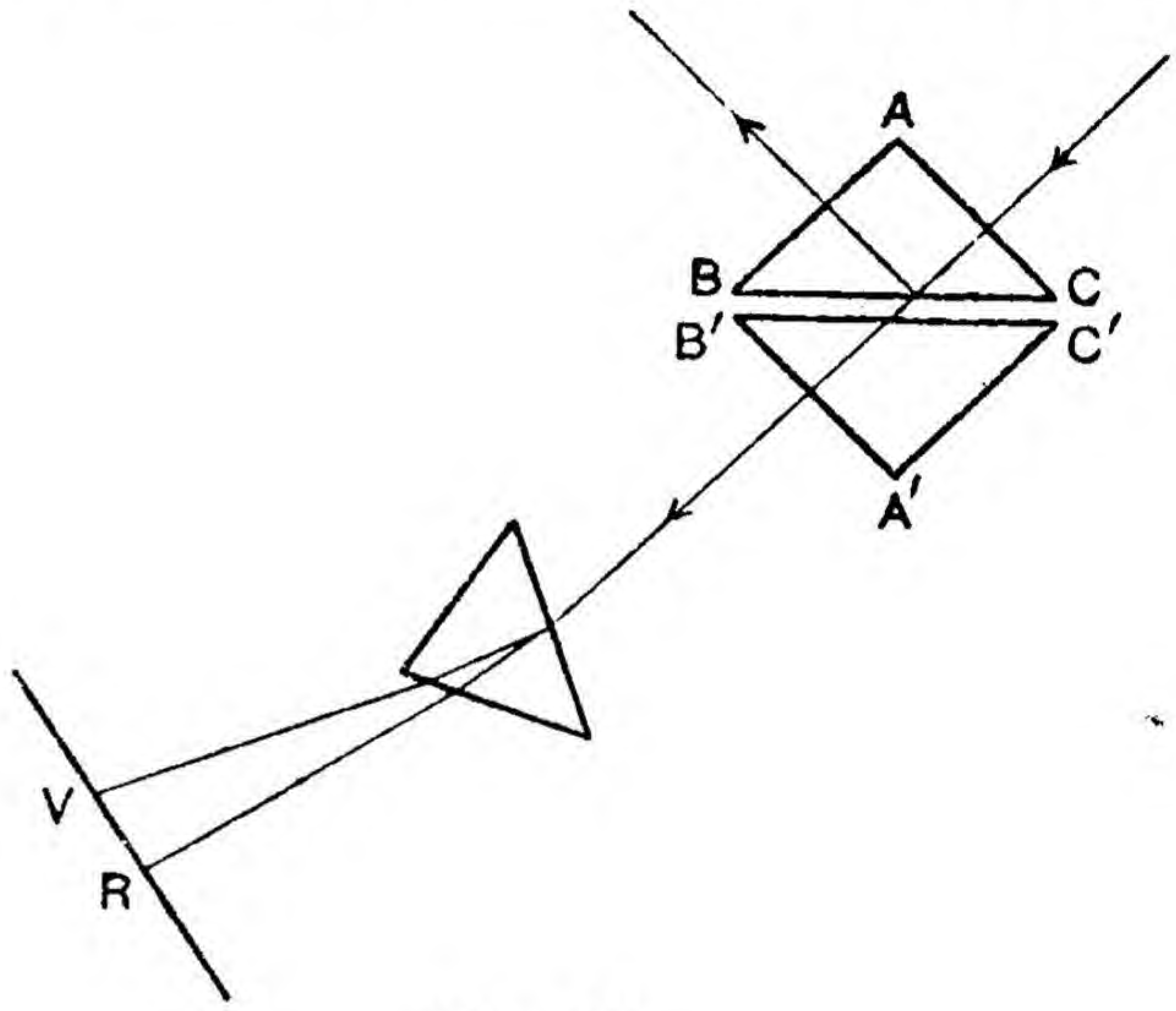


Fig. 50.

and after emerging from the prism a spectrum of this light was formed in the usual way. But most of the light went on through the base of the prism. The prism was then rotated in an anti-clockwise direction, when the violet light was the first to be totally reflected from the base BC and went to form part of the reflected beam emerging from the prism and the violet end of the spectrum of this light was strengthened. The first rays to be totally reflected had the greatest refractive index. As the prism was rotated further, rays of less and less refractive index were successively introduced into the reflected beam, as was shown by the intensification of less deviated parts of the spectrum, starting at the blue and going on through the green, yellow, orange, and red. So rays of different refractive index were produced by reflection and since the final reflected beam was white when all the light was totally reflected, this proved that white light consists of rays of varying refractive index. The second experiment was Newton's method of filtering rays of different refractive index out of white light, his filter being total reflection. Two isosceles right-angled prisms ABC and A′B′C′ were placed with their faces BC and B′C′ in contact and a beam of white light was sent normally on to the face AC of

the combination (Fig. 50). It passed straight through the two prisms, a little being lost by reflection at the faces BC and B′C′, and a spectrum was formed of the transmitted light in the usual way. The pair of prisms was then rotated in an anti-clockwise direction and when the light was incident on BC at an angle greater than the critical angle for violet light, it was totally reflected and so filtered out of the transmitted beam. This was shown by its absence from the spectrum VR. The colour of the transmitted beam changed. As the prism was rotated further, the colours of the spectrum were filtered out in order of decreasing refractive index, as shown by the order in which they disappeared from the spectrum VR. Thus the transmitted beam, which was identical with the white light incident on AC, since it underwent a refraction from glass to air at BC and air to glass at B′C′ whose possible effects on the light would cancel one another out, consisted of rays of varying refractive index. We may sum up the conclusions to be drawn from this group of experiments in the following way: it is possible to obtain rays which are deviated through different angles when falling at the same angle on the same prism either by reflection from bodies of natural colours, or by total reflection, or by refraction; this different amount of deviation is not a diffusion or spreading out of all the light, since the rays which are deviated least are always so deviated, and therefore rays can be produced which differ permanently in refractive index; finally, since such rays can be obtained from white light, it follows that white light consists of rays of varying refractive index.

31. TO PRODUCE A PURE SPECTRUM

Newton was quite aware of the fact that he was not obtaining a complete separation of the rays of different refractive index except at the edges of his spectra, as the finite size of the images of the sun produced

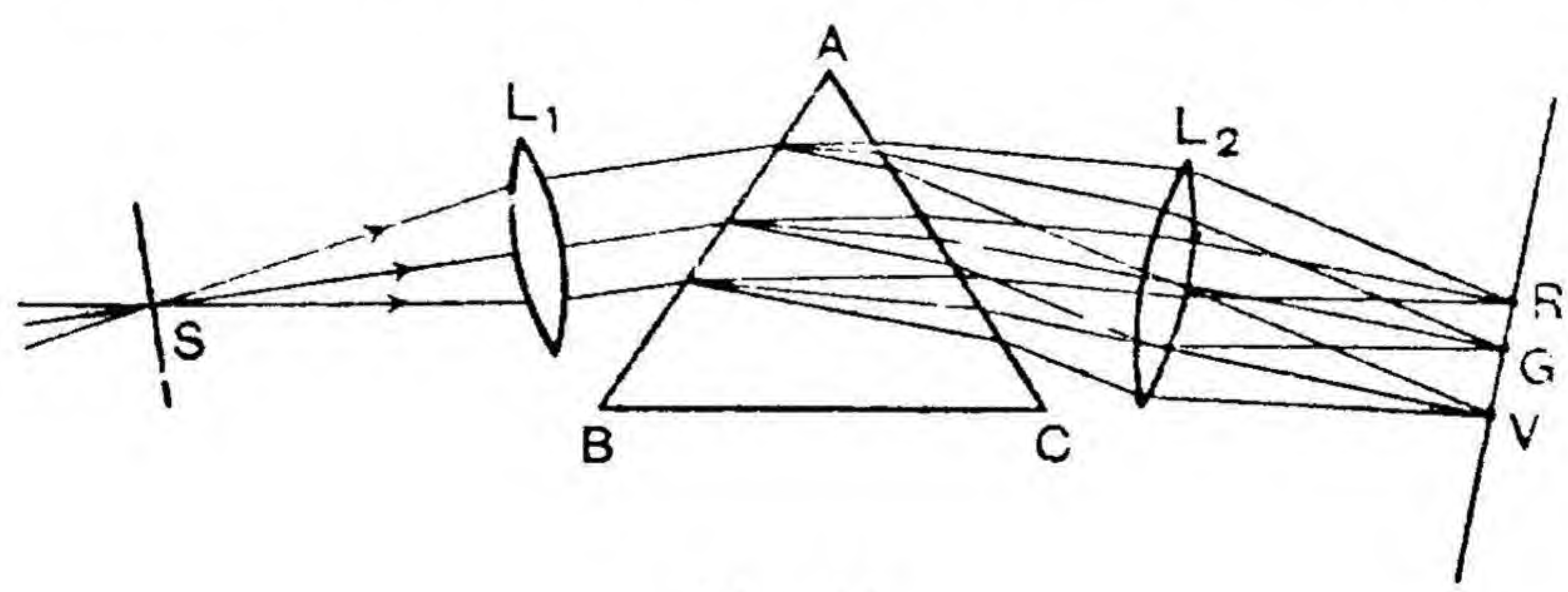


Fig. 51.

by the different sets of rays would cause overlapping everywhere else. He realised that it might be desirable to produce a spectrum free from this overlapping and consequent weakening of the various colours. Such a spectrum is called a pure spectrum, and Newton saw how to produce one. The method shown in Fig. 51 is not quite the one which he used, but it is the one most commonly used nowadays. A narrow slit, S,

whose length is perpendicular to the plane of the paper, is illuminated by some source of white light of finite size and is placed in the focal plane of the lens L_1. The rays from any one point of the slit are rendered parallel by the lens and fall on the prism ABC, which deviates each colour through a different angle, there being an emergent parallel beam corresponding to each colour. As each of these beams is travelling in a slightly different direction, they are each brought to a focus in a slightly different place in the focal plane of the lens L_2, where a screen is placed to receive the spectrum. The result is that a sharply focussed image of the slit in each different colour is formed on the screen, and as the slit and each image may be as small as $\frac{1}{2}$ mm., it will be seen that there is much less overlapping than in Newton's original experiment, where the size of the pinhole image of the sun was $2\frac{1}{2}$ in. This pure spectrum need not be any fainter than Newton's spectrum, since quite a widely divergent beam will emerge from the slit if a large enough source is used, and all these rays are collected by the two lenses and go to form the final images of the slit in the various colours. It is interesting as showing Newton's clear grasp of any problem that he suggested the use of a V-shaped slit, pointing out that a spectrum will be produced which will be very faint but very pure along the edge corresponding to the point of the slit and whose brightness will increase and purity decrease as we go down from one edge to the other in a direction at right angles to the length of the spectrum. Newton himself seems to have made very little use of the pure spectrum. The term **dispersion** is used to denote either the separation of white light into its constituent colours when it passes through a prism, or the variation of refractive index with colour which causes this separation.

32. COLOUR IS NOT A MIXTURE OF LIGHT AND DARK

Having established the facts that different colours have different refractive indices and that white light consists of rays of different refractive indices and therefore of different colours, Newton now proceeded to dispose of the view that colour is a mixture of light and darkness, the actual colour depending on the proportion of each constituent present. This idea probably arose because coloured light is usually less intense than white light and colours often occur at the edge of shadows, that is, at the meeting-place of light and darkness.

His first experiment consisted in producing a spectrum on a screen by sending a broad beam of white light on to a prism ABC (Fig. 52) and then passing the emergent beam through a small hole in a board on the side of the prism remote from the source of light, with the result that each colour emerged from the hole in a different direction and a fairly pure spectrum was formed on the screen. A narrow obstacle, L, was then introduced into the beam of white light incident on the prism. The reader should notice that the obstacle was introduced and the shadow

was produced before the white light was acted on by the prism. In this way the edge of the shadow could be produced at any point in the spectrum RV, and Newton showed that, at whatever point the shadow began or finished, the colour at that point remained the same. In other words, colour is not due to a mixture of light or darkness or to effects occurring at the termination of light and shade, for the colour at a given point of the spectrum remained the same however the position of the edge of the shadow was altered. He then introduced the obstacle so that

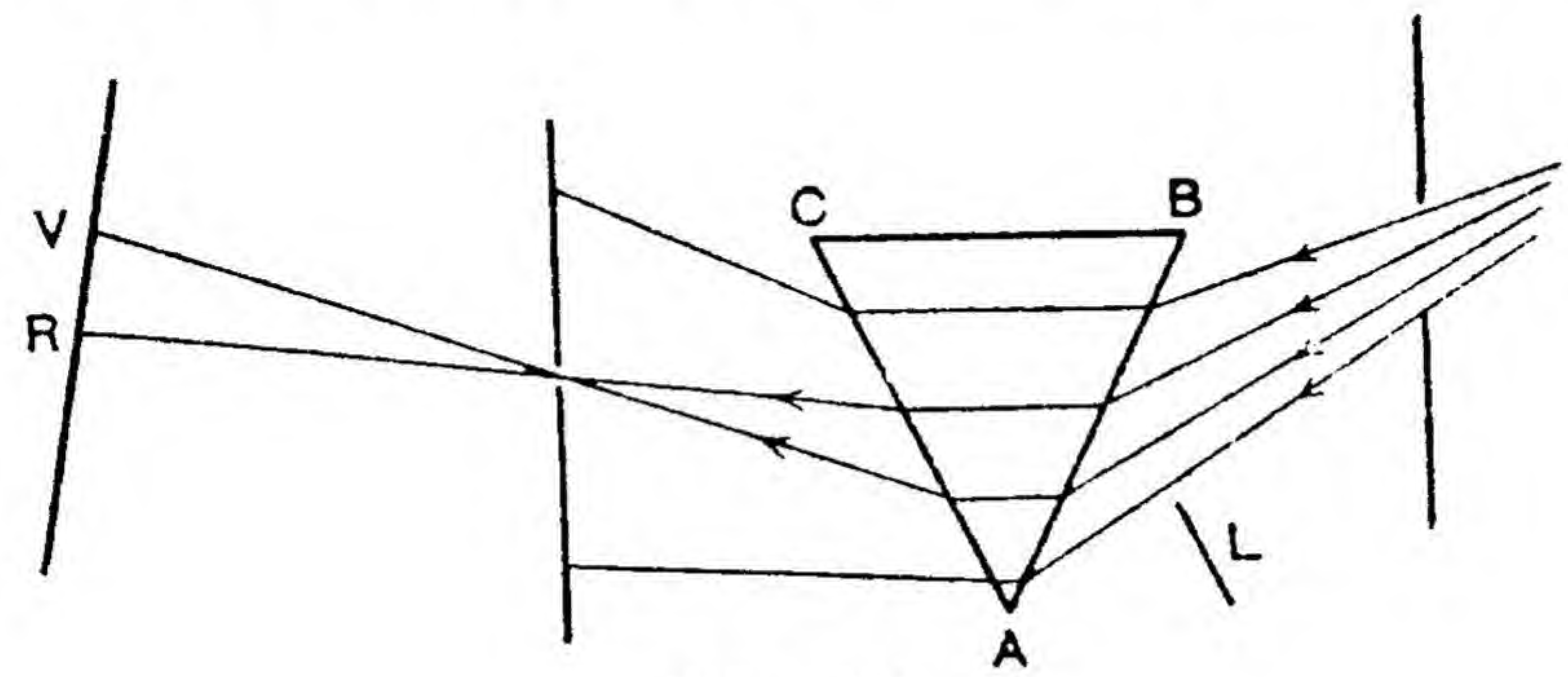


Fig. 52.

its length was normal to the plane of the diagram and put it right into the middle of the incident beam so that a shadow was produced right in the middle of the spectrum, and again no effect was produced on the colours immediately on either side of the shadow. Finally, he moved in two obstacles, such as L, at either side of the incident beam of white light, so that only a narrow beam of light could pass between them. Again the narrow band of colour remaining on the screen was the same as the colour existing in that place in the absence of the obstacles. Another important conclusion can be drawn from this experiment: the hole in

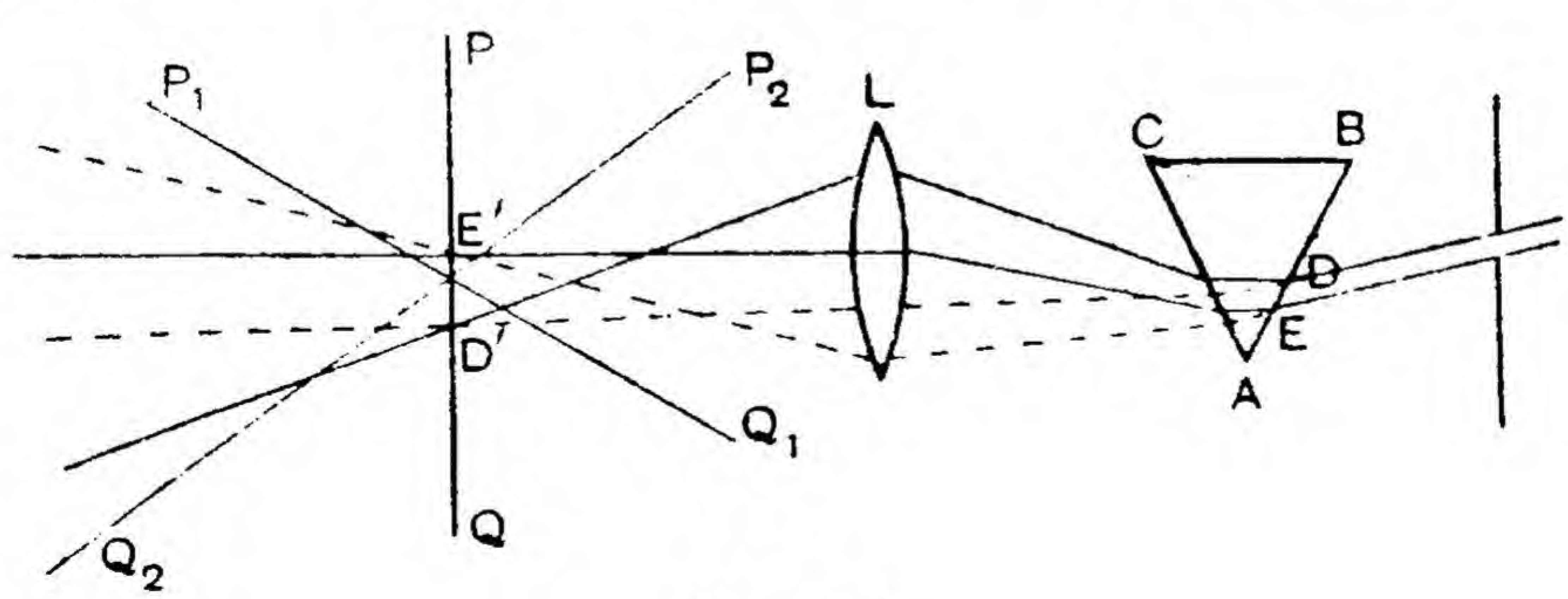


Fig. 53.

the screen was situated at a place where the light emerging from the prism was still white light, and yet a pure spectrum was produced at VR. This proves that the spectral colours are already present in the white light and that they are not manufactured by the refraction when they are produced in the usual way.

Newton further verified this conclusion in an interesting way by sending a narrow beam of white light on to a prism ABC, thus producing the usual spectrum on the lens L (Fig. 53). He arranged this lens so

that an image D′E′ of the patch of white light DE incident on the face AB of the prism was thrown on the screen PQ. This image was white, since the lens brought the various coloured rays which diverged from the point D to a focus at the point D′ again and so formed a white image of the point D. The same thing happened for every point of the object, and so a white image of the whole patch DE was produced at D′E′. But if the screen was tilted into the position P_1Q_1, the whole image took on a yellowish hue, or if the screen was tilted to P_2Q_2, the image assumed a bluish tinge. These colours were produced without altering in any way the edges of the beam or the confines of light and darkness, thus showing that colour cannot be due to effects at such edges. The yellowish tinge seen in the position P_1Q_1 was due to the fact that the red and yellow rays being the least deviated struck the screen more obliquely than the remaining colours, and so were reflected the more strongly, the converse being the case when the screen is at P_2Q_2.

33. LIGHT OF A DEFINITE REFRACTIVE INDEX HAS A DEFINITE COLOUR CORRESPONDING TO THAT REFRACTIVE INDEX AND THAT COLOUR CANNOT BE CHANGED BY REFRACTION AND REFLECTION

Newton was now able to take a very important step in his investigation of colour in that **he could replace a physiological sensation by a definite physical quantity, refractive index.** A subject has hardly reached the stage of a science until this has been done. For example, the study of heat hardly made any serious progress until the physiological sensation of hotness was replaced by the scientific quantity temperature. It is true that the idea of temperature is ultimately based on a sensation, but the more advanced scientific way of specifying a temperature leads to the possibility of dealing with a far wider range of temperature in a far more precise way, and it ultimately leads to the scientific *conception* of temperature as the average speed of the molecules of matter. In just the same way we can hope to make much more rapid progress in the study of light, if we can replace the sensation of colour by the physical quantity refractive index, which is expressed in numbers. Newton proved this proposition by forming a spectrum in his usual way by sending sunlight through a narrow hole in a blind and he cast the spectrum on to a piece of black paper with a small circular hole in it $\frac{1}{6}$ in. in diameter. If one particular colour was allowed to fall on the hole and was sent on to another prism, it was deviated by the second prism to the same extent as the first one if the conditions were identical and no new colour was produced by this second refraction. This showed that neither the refractive index nor the colour of the light was changed by the second refraction and the same thing was true of any number of refractions. The light of a definite colour and refractive index derived from a narrow portion of a pure spectrum is called **homogeneous light** or **mono-**

chromatic light or a **pure colour.** Newton verified this proof by viewing through a prism the hole in the black paper on which the spectrum was cast. As he expected, the image remained circular and the colour was unchanged, whatever colour was emerging from the hole. It follows therefore that the colours of the spectrum are the only homogeneous colours which can be obtained from white light.

Newton also showed that the colour of homogeneous light is not altered by reflection whether from polished mirrors, or from white paper, or from coloured bodies, although in this last case there is usually a decrease in intensity of the light. For example, red light is reflected with intensity undiminished from a red paint, but with greatly decreased intensity from a blue one.

Newton was well aware that the converse of the above proposition is not necessarily true. That is, light of a definite colour may not have a definite refractive index, although its colour is the same as that of a homogeneous light. This is due to the fact that such a colour may be produced by adding a number of homogeneous colours. He showed this by illuminating a white circle with homogeneous red and green light, the result being the same colour as homogeneous yellow. But if the circle were viewed through a prism it split into two images, one red and the other green, in the places where they did not overlap. Newton saw the bearing of this on the blurred images produced by telescopes, for he viewed flies, print, and other small objects through a prism. When they were illuminated with homogeneous light, they were sharply focussed, but when they were illuminated with white light, they were both blurred and coloured. The experiments in this article are important in that they form yet another refutation of Grimaldi's suggestion that rays of light, of whatever colour, are diffused or spread out when they pass through a prism. This is shown to be quite untrue for homogeneous light.

34. AN ATTEMPT TO FIX NATURAL COLOURS

As we shall see below, Newton realised that the blurred images produced by telescopes of large magnifying power were due to different colours having different refractive indices. If this defect is to be removed, it will be necessary to find substances which separate two given colours to the same extent while producing different deviations of the spectrum as a whole. So it was natural that an attempt should be made to fix definite colours in the spectrum, although it appeared to show a continuous variation from red through green to violet. In doing this Newton used a method which is of common application in Science; he used the known to help him to interpret the unknown. It is known that the lengths of a vibrating string sounding the notes of the diatonic scale are in the ratio of 1, $\frac{8}{9}$, $\frac{5}{6}$, $\frac{3}{4}$, $\frac{2}{3}$, $\frac{3}{5}$, $\frac{9}{16}$, $\frac{1}{2}$. Newton asked an assistant, who was more sensitive to colour than himself, to divide the spectrum into its constituent colours, and he then found that the lengths of the spectrum

measured from the origin O (Fig. 54), which is twice as far from the violet end as the red, were in the same ratio as the lengths of a vibrating string sounding the notes of the diatonic scale. It was quite natural, in view of the limited range of material at his disposal and the fact that no other way of producing a spectrum such as the diffraction grating had yet been discovered, that Newton should attach some real significance

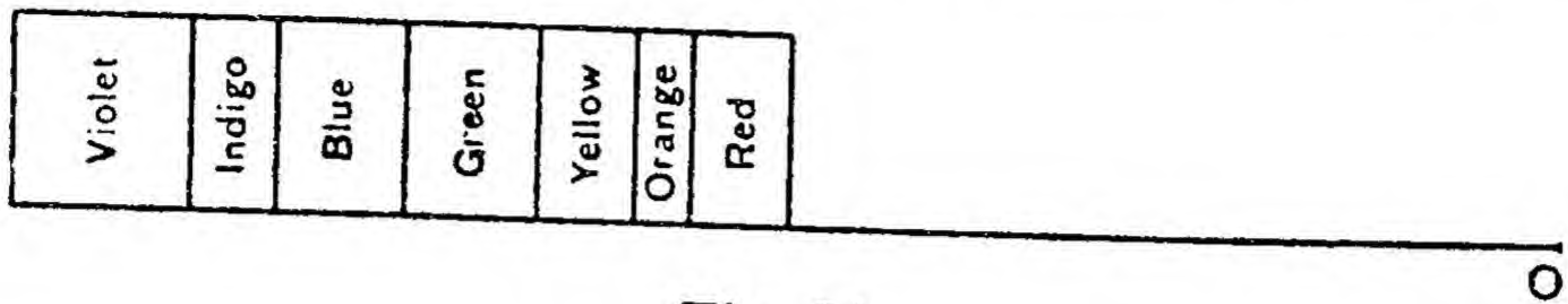


Fig. 54.

to this result. In view of the much greater body of evidence now at our disposal, we know that there is no theoretical significance in this subdivision of the colours of the spectrum, but it was quite a sound way of trying to fix different colours at that time with a view to comparing the dispersions of different materials.

35. THE COLOURS OF NATURAL OBJECTS AND LIQUIDS

Newton proceeded to give a clear explanation of the colours of solid bodies and of liquids. A red body looks red because it reflects only red light, or reflects this colour much more copiously than any other. This view is not quite correct, since we know now that a slab of smooth sealing-wax, for example, reflects white light regularly without producing any colour. It is more true to say, as Newton did in explaining the colour of gold both in bulk and in thin films, that a red body looks red because the white light penetrates a thin surface layer before being reflected, and this surface layer absorbs most strongly all colours but red. Thus we see the difference between the old and new views. On the former, a body looks red because it tinges the white light with the red colour, as it were ; it gives something *to* the white light. On the latter, it takes something *away from* the white light to reveal the red colour which is in the white light all the time. This view also explains why a red body looks a bright red if viewed in homogeneous red light; whereas it looks a very faint blue if illuminated with homogeneous blue light. In the second case, it absorbs very nearly all of the kind of light falling on it and so looks faint.

A blue liquid looks blue in the same way because it absorbs every colour except blue, as can easily be seen by forming a spectrum in the usual way and putting a tank of the liquid somewhere between the source and screen. Every colour in the spectrum will be missing except the blue. It is now possible to buy special filters, known as Wratten filters, which absorb all the spectrum except a very narrow region. Gold looks yellow because it reflects a narrow region round and about the yellow ;

there is no question here of the light entering a very thin surface layer and all the colours but yellow being absorbed ; the process is one of selective *reflection* and the remaining colours are absorbed in the body of the gold. But if a *thin* film of gold is viewed in transmitted light it looks blue-green, the complementary colour to yellow, since the absorption of the colours other than yellow is only partial with thin films.

36. COLOUR BY ADDITION

We have already seen that to each refractive index there corresponds a definite colour, called a homogeneous colour, which cannot be changed by reflection, refraction, or any other process. But the converse proposition is not necessarily true, as was well known to Newton. This is due to the fact that when two homogeneous colours are added, a third homogeneous colour may be produced. For example, if two circular patches of homogeneous red and green light are cast on a white screen and made to overlap, the overlap will appear the same colour as homogeneous yellow, if the intensities of the two patches are properly adjusted. So a new colour is made by allowing the eye to be acted on by two homogeneous colours simultaneously. This should be contrasted with the result of mixing two pigments or two filters ; if a red and green filter are placed in a beam of white light, the resulting colour is black because the red filter absorbs every colour but red and the green one every colour but green, and so the two together absorb every colour. If red and green pigments are mixed, the result is a dirty brown. Again, if yellow and blue patches are cast on a screen, the colour in the overlap is white, if the intensities are suitably adjusted. But yellow and blue pigments mixed give green, the only colour which neither pigment absorbs. These results are quite in accordance with Newton's view of colour, but they cannot be explained on any chemical theory. Finally, the addition of two homogeneous colours may produce a colour which is not to be found in the spectrum at all. For example, if pure red and violet are added the result is purple, which does not appear in the spectrum. Newton was able to suggest a rule from which the colour produced by mixing two given homogeneous colours can be predicted when the intensity of the colours is given.

The fact that a given pure colour, such as yellow, can be produced either by spectral yellow or by the addition of two or more pure colours of suitable intensities suggests that the mechanism by which the eye detects colour is very different from that by which the ear detects pitch. It requires a source of frequency 256 c.p.s. to produce the pitch known as middle C and a note of frequency 258 c.p.s. will not be accepted by the ear as middle C, nor will any combination of pure notes be accepted, as the ear will detect the separate components. This problem cannot conveniently be discussed now, and so it will be postponed until the eye is fully treated in Chapter 7.

37. THE SYNTHESIS OF WHITE LIGHT

There is only one link needed to complete the chain of evidence in favour of Newton's view of colour, and that is the synthesis of white light from the spectral colours and the proof that, in making white light, those colours do not react with one another in any way. Newton did this by forming a spectrum in the usual way by sending sunlight through a small hole in a screen on to a prism P_1 and so forming a spectrum on the lens L (Fig. 55). This lens was placed so that it formed a real image A_2B_2 of the patch A_1B_1 of the prism P_1 illuminated by the incident light. A second prism P_2 was placed so that A_2B_2 falls on its second face and it was turned so that all the rays were rendered parallel after emerging from it, and so all the spectral colours were recombined after emerging from this prism. For the ray of white light striking A_1 was split up into the spectral colours by the first prism and these colours were all brought to a focus at the point A_2 by the lens L. Their directions were all different though, and so they were passed through a second prism, which bent the

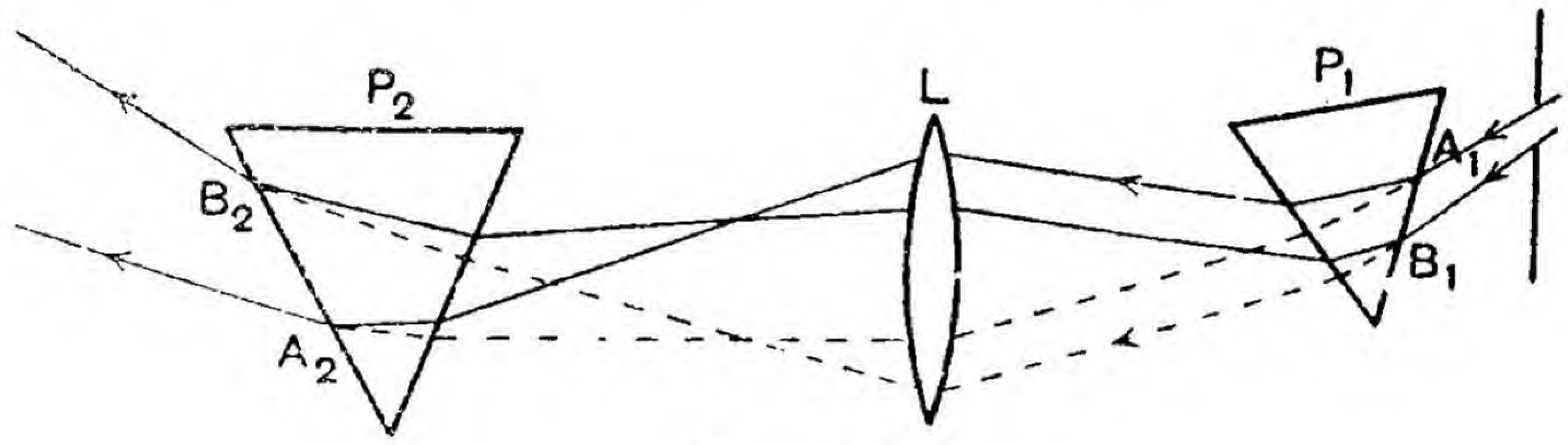


Fig. 55.

violet more than the red and made the directions of all the rays the same, while keeping them coincident. Hence the rays from any point within A_1B_1 emerged from the second prism at the same point and in the same direction, and so a beam of light was made which was composed of the spectral colours. Some little adjustment of the distance of the second prism from the lens was needed to get a beam of white light untainted by any colour. This beam looked the same as sunlight, and all the experiments which had been performed with such light could be performed with this artificial sunlight and yielded the same result. Finally, it was shown that the spectral colours do not react in any way with one another when making white light, but maintain their independence, by placing a screen at L so as to cut out, say, the red. The colour of the synthesised beam changed, of course, but the important thing is that, if the beam emerging from the second prism was analysed with a prism, its spectrum was found to be deficient in red. This was true of any and every colour in the spectrum and so these colours maintain their independence, even when they are all present together producing the effect of white light.

Thus Newton established by experiment and observation that colour is not something given to white light by the body, from which it is reflected

or through which it has passed ; it is not a mixture of light and darkness, nor is it a kind of chemical compound of two or three elementary colours. He had proved that white light consists of a set of independent pure colours, each characterised by a refractive index, which is invariable for a given substance and which cannot be changed by reflections or refractions. These colours do not act on one another in any way when producing white light, which is due to their simultaneous action on the eye. The colour of pigments and liquids is due to their absorbing one or more of these pure colours from the incident white light, the remaining colours adding either to another pure colour or to some colour which is not in the spectrum.

38. CHROMATIC ABERRATION AND NEWTON'S ATTEMPT TO ELIMINATE IT

The cause of the blurred images in telescopes is now quite apparent. We have already seen that a lens brings blue light to a focus at a different place from red light and so, if a beam of white light is incident on a lens,

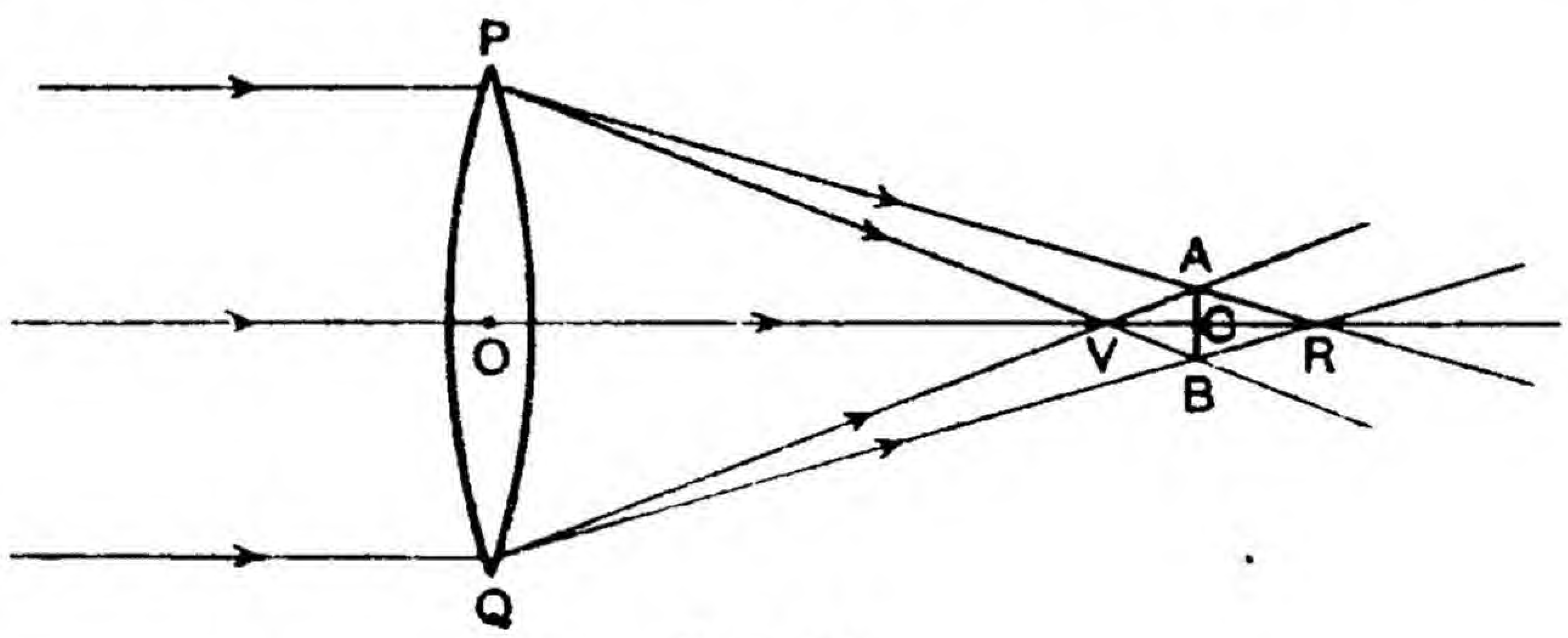


Fig. 56.

each colour will come to a focus at a slightly different place and a blurred image will result. For example, if a beam of white light parallel to the axis of the lens PQ falls on the lens (Fig. 56), the violet rays will come to a focus at V, while the red rays, being less refrangible, will come to a focus somewhat further from the lens at R, the other colours each having their own foci at intermediate points. So a violet point in the centre of a red ring will be seen on a screen at V and the converse on a screen at R. Also the best focus which is obtained from this beam, which comes from a point object at infinity, is the circle of least confusion ACB formed by the intersection of the rays diverging from V with those converging on R. The same is true of a point object at a finite distance from the lens and also of the collection of point objects into which any object of finite size may be analysed, and so the images of all such objects formed by a lens will be blurred and usually coloured to some extent. This defect in the image formed by a lens due to the variation of refractive index with colour is called **chromatic aberration**, and its existence can be proved in another way which enables an estimate of the diameter

of the circle of least confusion to be made. The focal length of a lens is given by equation (11)

$$\frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

Considering only the case of lenses in air, $n_1=1$ and n_2 may be written as n and denotes the refractive index of the material of the lens for sodium yellow light. Calling the refractive index for violet and red light n_v and n_r respectively, the focal lengths f_v and f_r of the lens for the same two colours respectively is given by

$$\frac{1}{f_v}=(n_v-1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

$$\frac{1}{f_r}=(n_r-1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

Since $n_v>n_r$, then $f_v<f_r$, which is just what has been found experimentally. Also

$$\frac{1}{f_v}-\frac{1}{f_r}=(n_v-n_r)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

$$\therefore\ f_r-f_v=(n_v-n_r)\left(\frac{1}{r_1}-\frac{1}{r_2}\right) f_r\, f_v$$

Putting $f_r=f_v=f$ and $\left(\frac{1}{r_1}-\frac{1}{r_2}\right)=\frac{1}{f(n-1)}$ we have

$$f_r-f_v=\left(\frac{n_v-n_r}{n-1}\right)\cdot f \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (20)$$

Now $$\frac{\mathrm{AB}}{\mathrm{PQ}}=\frac{\mathrm{CR}}{\mathrm{OR}} \text{ (Fig. 56).}$$

Assuming that C bisects VR, which is nearly true,

$$\mathrm{AB}=\frac{(n_v-n_r)}{2(n-1)}\cdot\frac{f}{f_r}\cdot\mathrm{PQ}$$

Putting $f_r=f$, we have the following approximate expression for the diameter of the circle of least confusion :

$$\mathrm{AB}=\left(\frac{n_v-n_r}{n-1}\right)\cdot\frac{\mathrm{PQ}}{2} \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (21)$$

So the diameter of the circle of least confusion is determined by the diameter of the lens and by the quantity $\frac{n_v-n_r}{n-1}$, which is called the **dispersive power** of the material of the lens. The existence of the circle of least confusion is due to the variation of refractive index with colour, which is known as dispersion, and the expression $\frac{n_v-n_r}{n-1}$ measures

this property of the medium in numbers. Taking a typical case, a telescopic objective of crown glass for which PQ=20 cm., $n=1{\cdot}55$, $n_v=1{\cdot}56$, $n_r=1{\cdot}54$,

$$AB=\frac{0{\cdot}02}{0{\cdot}55}\cdot\frac{20}{2}=0{\cdot}36 \text{ cm.}$$

The circle of least confusion due to spherical aberration is only 0·032 cm. in diameter (Art. 45) and so the blurred images of telescopes is due largely to chromatic aberration. In view of this result, it is surprising that they are not more blurred. But, as Newton pointed out, the above refractive indices refer to the extreme limits of the spectrum, whereas the eye is most sensitive to a much narrower region bounded by the orange on one side and the blue-green on the other, and this will make the diameter of the circle perceived by the eye smaller and will account for an amount of blurring consistent with what is observed in practice.

This simple analysis of the cause of chromatic aberration suggests a way of eliminating it. The same defect will occur in a diverging lens, which will have the focus for violet rays nearer to the lens than that for red rays. So, if the converging lens is combined with a suitable diverging lens, since their aberrations are in opposite directions, they may compensate out each other. Equation (20) shows that, if the dispersive power of the materials of the two lenses is the same, the numerical value of their focal lengths will also be equal. In that case they will be equivalent to a parallel-sided glass plate and will cease to act as a lens ; in other words, in eliminating the dispersion, the deviation has also been destroyed. A combination of lenses, which has the same focal length for all colours and is therefore free from chromatic aberration, is called an **achromatic lens**, and the only way to produce one consisting of two lenses in contact is to find two materials of different dispersive power. This was a difficult thing for Newton to do, because of the difficulty of fixing definite colours in the spectrum. For example, if n_v and n_r are taken to be at the extremities of the spectrum, it is impossible to measure them accurately, since the limits of the spectrum are not accurately defined. It was possibly for this reason that Newton suggested the division of the spectrum into colours, which has been noticed above (Art. 34), but this is hardly any better, since the division cannot be made without fixing the limits of the spectrum first. At all events, Newton came to the conclusion that the dispersive power of all materials is the same. This statement is based on measurement and, in criticising it, it must be remembered that the measurements were difficult because of the absence of definite colours to work with and also there was a much smaller range of materials at Newton's disposal than is the case now. So Newton gave up the attempt to eliminate chromatic aberration from refracting telescopes and turned his attention to reflecting telescopes.

EXAMPLES ON CHAPTER IV

1. Trace clearly the steps by which Newton replaced the physiological sensation of colour by a physical quantity, refractive index. Discuss the bearing of this on the cause of the blurred and coloured images of telescopes.

2. Discuss Newton's experiments on the constitution of white light and state concisely the results he obtained.

3. Write a short account of experiments designed to show that white light may be considered to be a mixture of coloured lights.

Solutions of two dyes A and B are examined in a spectroscope. A transmits light in the spectral ranges 6500 Å.U. (red) to 4800 Å.U. (green), and B transmits in the ranges 5000 Å.U. (green) to 4000 Å.U. (blue). Describe and explain what you would expect to see with the naked eye in the following cases:

(a) White paper is dyed in a mixture of A and B and illuminated with white light.

(b) White paper is dyed in A and illuminated with light which has passed through B.

(c) White paper is illuminated simultaneously with two beams of light, one of which has passed through a solution of A and the other through a solution of B.

(d) White paper is illuminated with light which has passed in succession through a solution of A and a solution of B. *(Camb. Schol.)*

4. What is meant by dispersion in the case of light, and how is it related to the velocity of propagation of light in material media?

Light from a bright star is observed to be cut off by the dark limb of the moon. Discuss the effect of dispersion in the earth's atmosphere in producing possible colour effects at the instant of eclipse. How long will these effects last if the difference of the refractive indices of air of standard density for blue and red light is 8×10^{-6} and the atmosphere is regarded as of uniform standard density equal to 10^{-4} of that of mercury? *(Camb. Schol.)*

5. Explain three of the following:

(a) The colours seen round the edges of the images of objects viewed through a telescope made with ordinary (uncorrected) lenses.

(b) The colour of the setting sun.

(c) The change in the colour of objects when they are taken from daylight into artificial light.

(d) The similarity of the colours of the light emitted by a sodium flame and an amber traffic signal and the difference between their spectra. *(N.U.J.B.)*

6. A point object is placed on the axis of an equi-convex lens 30·0 cm. from its centre. The radius of curvature of each face is 20·0 cm. and the refractive index of the glass is 1·635 for red light and 1·650 for violet light. Find the position of the red image and the violet image and the approximate diameter of the circle of least confusion, if that of the circular outline of the lens is 6·3 cm.

Chapter V

THE MEASUREMENT OF REFRACTIVE INDEX; SPECTRA; THE PRODUCTION OF ACHROMATIC LENSES

39. THE SPECTROMETER

It is evident that the measurement of the refractive index of materials for different colours and the examination of spectra is going to be important in this matter of eliminating chromatic aberration, and the spectrometer is the instrument which is used for these purposes. The general law of refraction is given by the equation

$$n_1 \sin i_1 = n_2 \sin i_2$$

but, since the refractive index relative to air is usually measured, we shall put $n_1 = 1$ and write n_2 as n, the refractive index of the material relative to air, when the equation becomes

$$\sin i_1 = n \sin i_2$$

So the measurement of the refractive index of a material involves measuring the angles of inclination of light of a given colour both in air and the material. This could be done directly using a slab of the material, but, in the spectrometer, a prism is used, and it is found more accurate to measure two other angles, the angle of the prism and the angle of minimum deviation of the given colour when passing through the prism, from which the above two angles, and hence the refractive index, can be deduced in the following way.

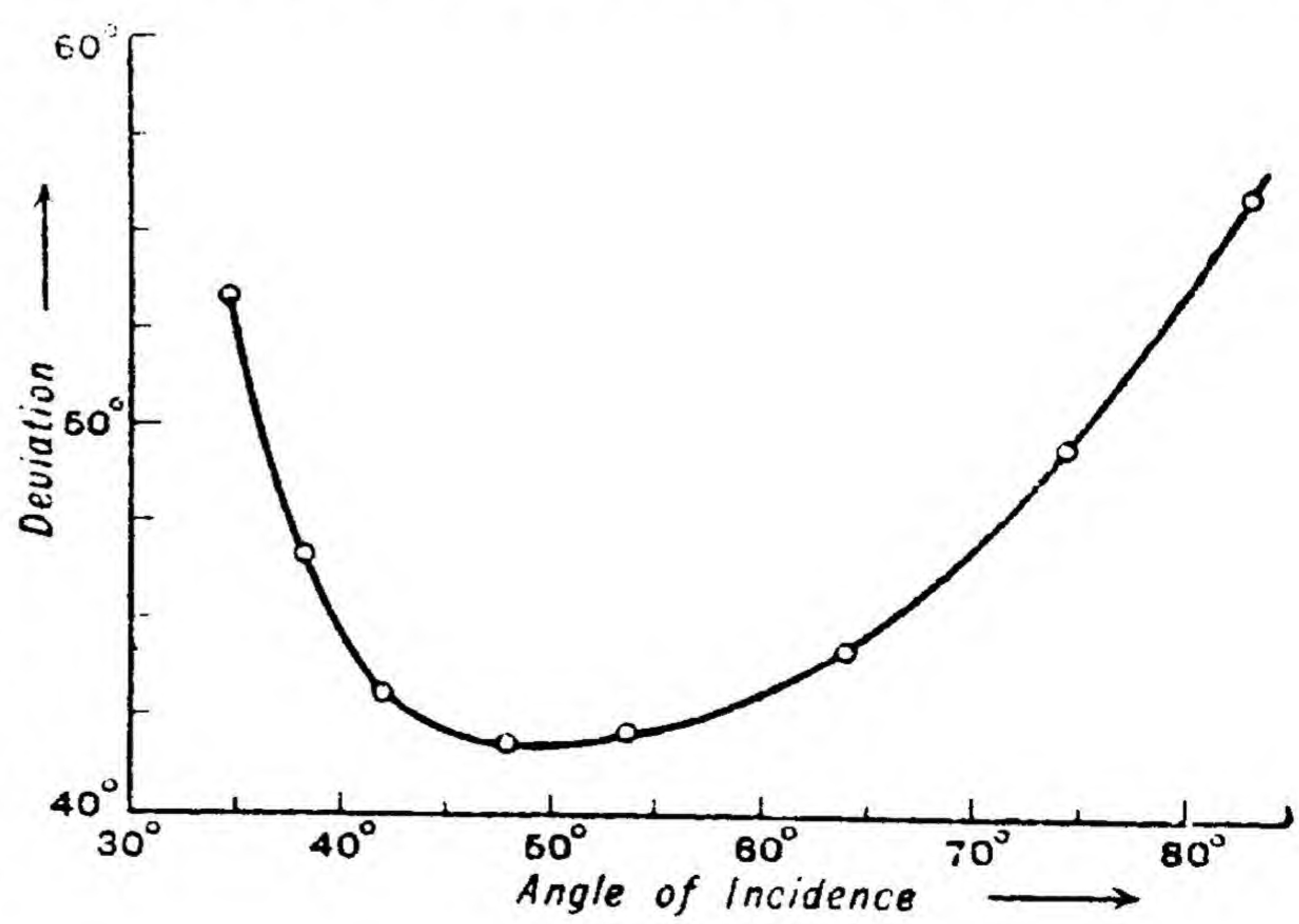


Fig. 57.

If a ray of light is sent through a prism and its deviation is measured for various values of the angle of inclination in air at the first face of the prism, the graph of deviation and angle of inclination takes the form shown in Fig. 57. There is one angle of inclination for which the deviation is

a minimum. It follows from the principle of the reversibility of rays of light, that the ray of minimum deviation passes through the prism symmetrically, so that the angle at which it emerges from the prism is equal to the angle at which it enters. For if it did not, and the ray of light were reversed, it would retrace its path and still pass through the prism at minimum deviation but with a different angle of inclination, which is contrary to experience. Let EFGH represent a ray of light of refractive index n for the material of the prism passing at minimum deviation through a prism PQR of angle A (Fig. 58). If i_1 and i_2 are the

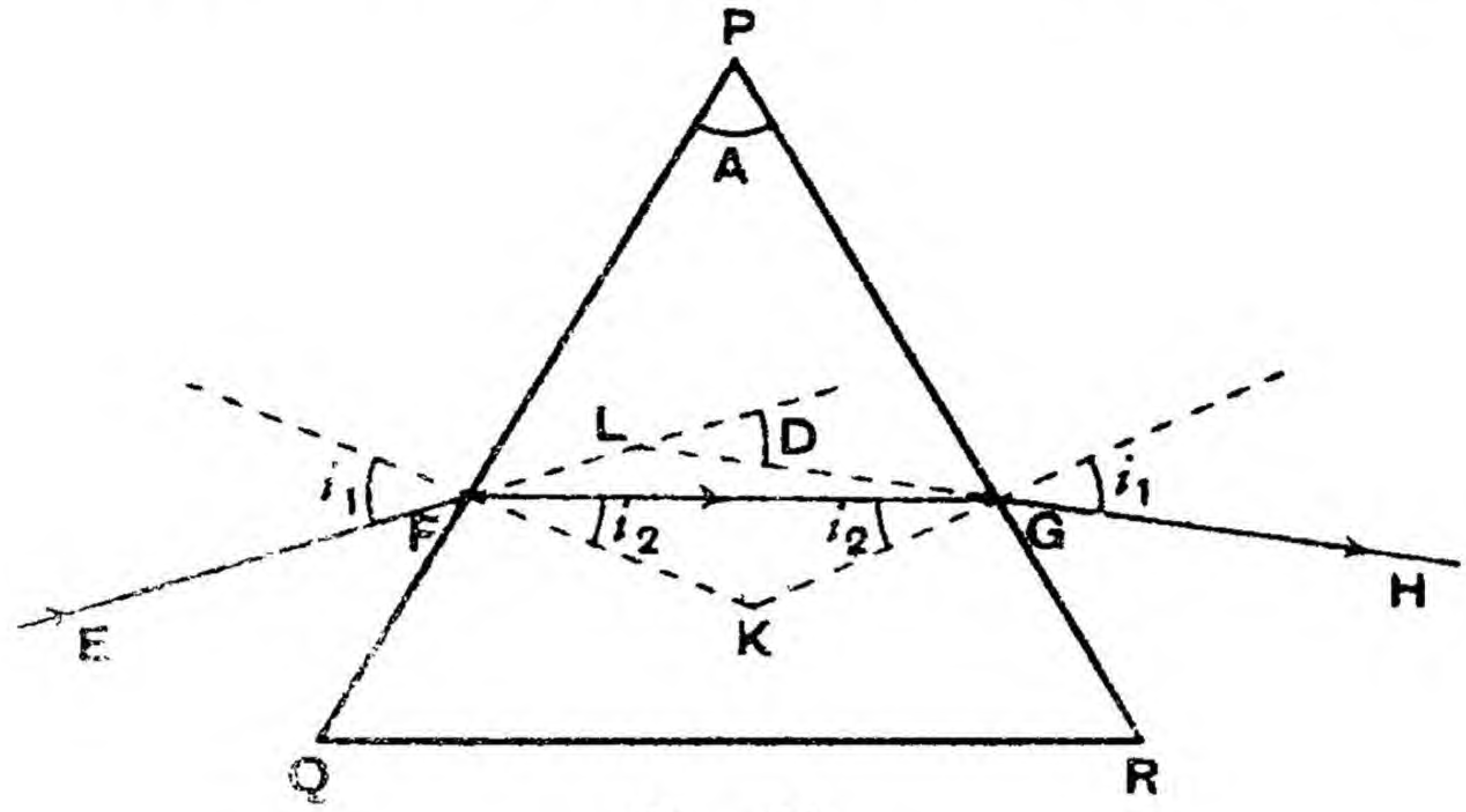


Fig. 58.

angles of inclination in air and the prism respectively, the angle of minimum deviation D is given by

$$D = \angle L\hat{F}G + \angle L\hat{G}F$$
$$= (i_1 - i_2) + (i_1 - i_2)$$
$$= 2(i_1 - i_2)$$
$$\therefore \quad D = 2i_1 - 2i_2$$

But
$$2i_2 = \pi - \angle F\hat{K}G$$
$$= \text{A, since PFKG is a cyclic quadrilateral.}$$
$$\therefore \quad i_2 = \frac{A}{2} \text{ and } i_1 = \frac{D+A}{2}$$
$$\therefore \quad n = \frac{\sin i_1}{\sin i_2} = \frac{\sin \frac{D+A}{2}}{\sin \frac{A}{2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

So the measurement of the angle of the prism and the angle of minimum deviation, at which the light of given colour passes through it, serves to give the angles of inclination of the ray in air and the material of the prism, and hence its refractive index for that colour.

The spectrometer is an instrument designed for the purpose of enabling these and other such angles, which may be needed in measurements on

spectra, to be found. It is essentially an instrument for producing spectra and measuring angles. It consists of a telescope T, a turntable B, and a collimator C, which is mounted on a base which rests on the bench (Fig. 59). The turntable and telescope rotate about a common axis passing through the centre of the turntable normal to its plane and both the axis of the telescope and the collimator pass through this axis of rotation and are at right angles to it. A circular scale L marked off in degrees is attached to the telescope and verniers V reading on this scale are attached to the turntable, and so the angle through which either the telescope or the turntable has been turned, while the other has been kept fixed, can be measured. The source of light whose spectrum is to be examined, say ordinary white light, is placed so as to illuminate the slit of the collimator, which is in the focal plane of the collimator lens. The light from any one point of the slit emerges from the lens as a parallel beam and is refracted by the prism P standing on the turntable B so as to produce a parallel beam for each colour travelling in a slightly different direction.

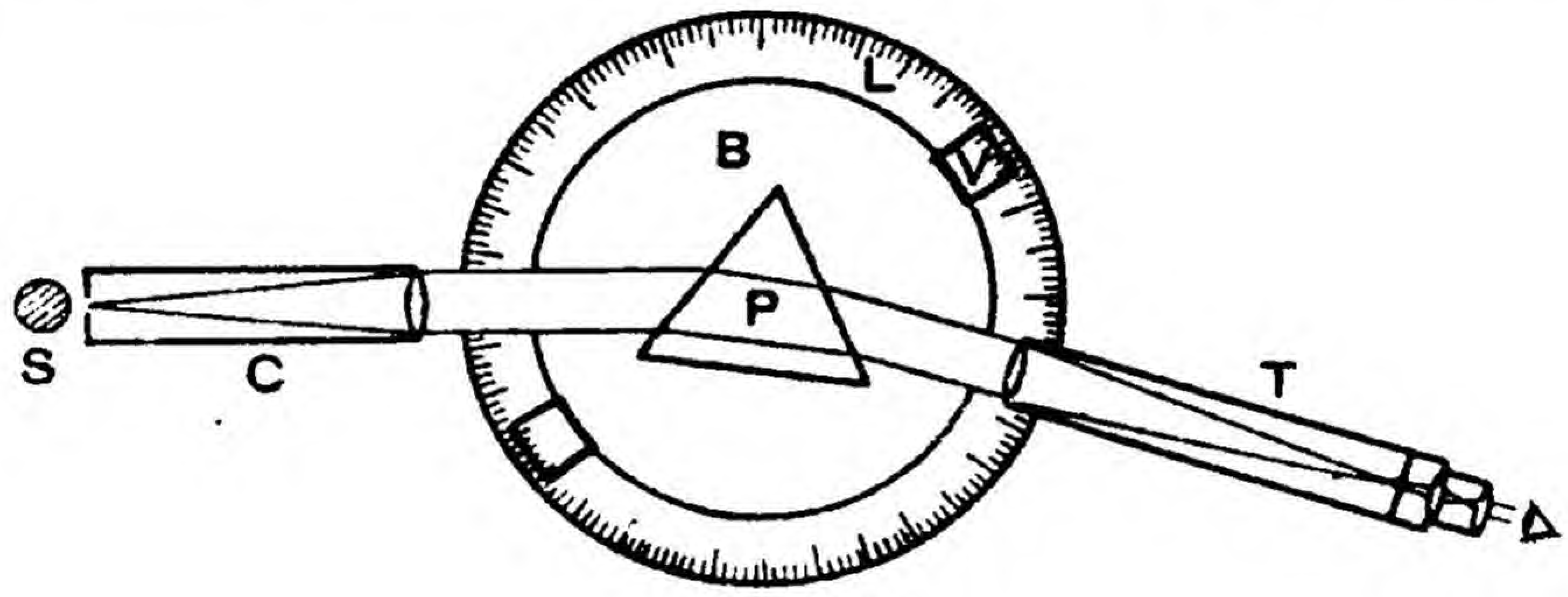


Fig. 59.

Instead of casting the spectrum on a screen, it is observed through the telescope T, which is turned so as to receive the light emerging from the prism, and each parallel beam is brought to a focus at a different point in the focal plane of the objective. This is done for every point of the slit and so a set of images of the slit in all the colours of the spectrum are produced in the focal plane of the objective. That is, the spectrometer forms a pure spectrum in the focal plane of the telescope objective, which is then observed through the eyepiece, whose function is to magnify the spectrum and enable more detail to be seen than is possible with the unaided eye.

Let us imagine that the refractive index of a given material for sodium yellow light is to be measured. Before this can be done, a number of adjustments have to be made to the spectrometer.

(*a*) Turn the telescope towards a well-illuminated wall near to it and focus the cross-wires by moving the eyepiece relative to them.

(*b*) Move the cross wires and eyepiece relative to the objective, so that distant objects are sharply in focus and coincide with the cross wires without parallax. It is essential to view the objects without a plate-glass window between them and the spectrometer, otherwise it is impossible

to get a sharp focus. This adjustment ensures that parallel rays entering the telescope come to a focus at the cross wires.

(*c*) The slit is now illuminated with sodium yellow light from a sodium lamp and the telescope is turned to point towards the collimator, so that an image of the slit can be seen. It will not, in general, be in focus. The slit is then moved relative to the collimator lens, until it is sharply in focus. Since the telescope has been adjusted, so that parallel rays come to a focus at the cross wires, it follows that the rays from any one point of the slit are emerging from the collimator lens as a parallel beam and are therefore coming to a focus at the cross wires, and therefore the observer sees the slit in focus, since the eyepiece has been focussed on the cross wires.

The spectrometer is now ready for use and a prism of the given material is placed on the turntable with its refracting edge normal to the plane of the table. The telescope is turned, until the refracted light is received, when an image of the slit will be seen, and the prism is turned by rotating the turntable, until the image reverses its direction of rotation. The setting at which this reversal occurs is the position of minimum deviation, and, when it has been found, the telescope is turned so that the vertical cross wire coincides with the centre of the image of the slit as it reverses, the slit being made as narrow as is consistent with being able to see its image. The reading of the vernier V on the scale L is taken and the prism is turned round, while keeping the turntable fixed, so as to deviate the light to the opposite side of the collimator and the reading of the telescope is taken for this second position of minimum deviation. Half the difference between these two readings is the angle of minimum deviation.

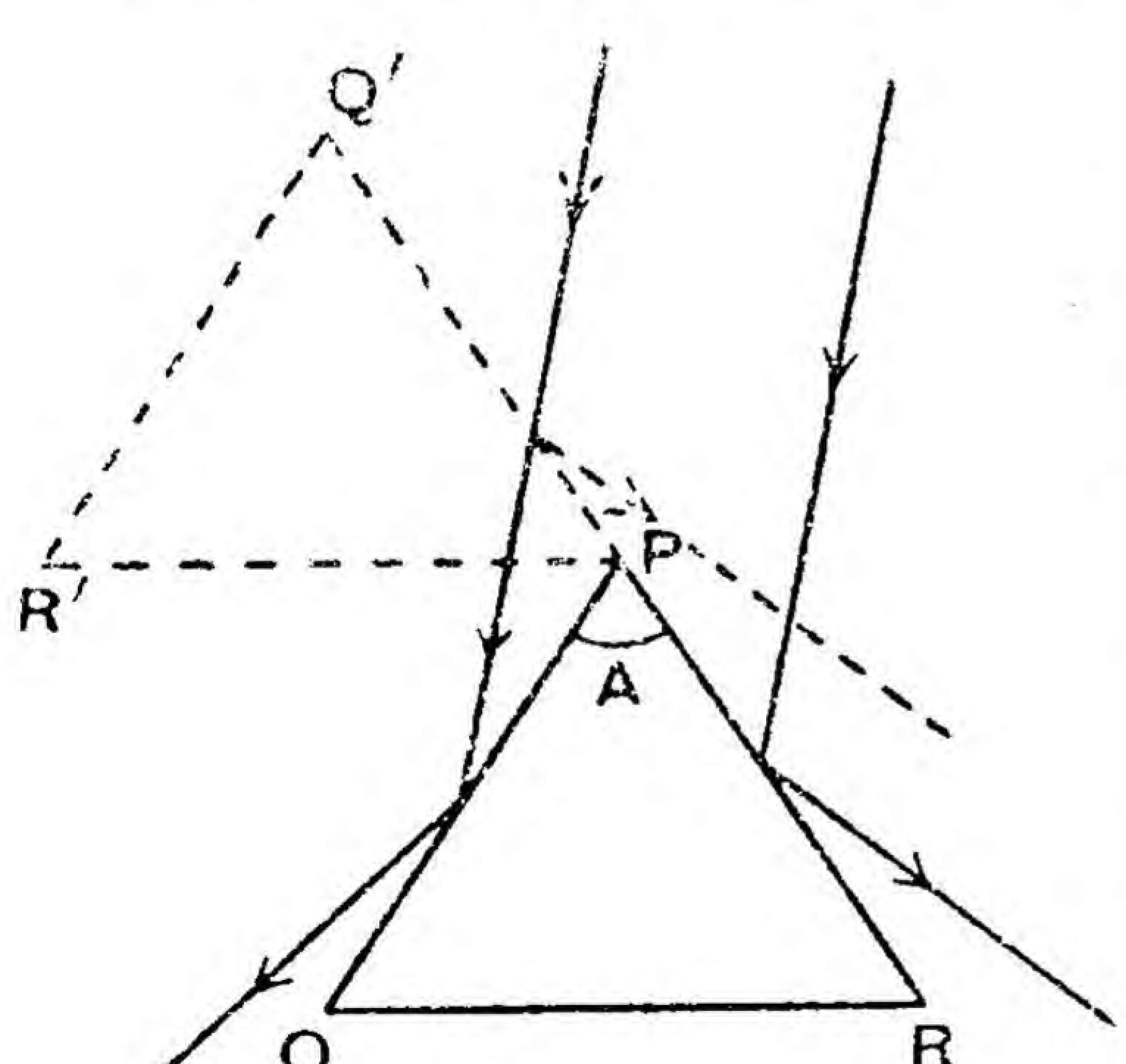


Fig. 60.

The angle of the prism is found by turning it so that the refracting edge points towards the collimator and turning the telescope to receive the beam of light reflected from one face of the prism. A faint image of the slit by reflection in one face of the prism will be seen and the telescope is moved, until the vertical cross wire coincides with this image, and the reading of the vernier is taken. The telescope is then rotated, until the vertical cross wire coincides with the image of the slit produced by light reflected from the other face of the prism, and the vernier reading is again taken. The angle of the prism is half the difference between these two readings. For consider a ray reflected from the face PQ of the prism PQR (Fig. 60). If the prism is rotated into the position PQ′R′, so that RPQ′ is a straight line, the ray reflected from PQ′ will be parallel to that originally reflected from PR. Since the reflecting surface PQ has been rotated through $180° - A$, the

reflected ray has been turned through $360°-2A$, and therefore the least angle between the rays reflected from the faces PQ and PR is $2A$, which is the difference in the two readings of the telescope. By substituting the values of the angle of minimum deviation and the angle of the prism in equation (22), the angles of inclination of the least deviated ray in air and the given material can be obtained and so the refractive index of the given material for sodium yellow light can be calculated.

40. CONTINUOUS, LINE, AND BAND SPECTRA

Newton found it difficult to measure the dispersive power of materials because of the absence of definite colours in the spectrum, but the reader will be familiar with the sodium yellow light, which has been mentioned in the previous article in connection with the measurement of refractive index, and it is now time to learn how Newton's difficulty was overcome. The fact is that there are three kinds of spectra: continuous spectra, line spectra, and band spectra. Newton was only familiar with the first of these, which is produced by any incandescent solid and consists of a continuous band of colour from the red to the violet end of the spectrum. A line spectrum is produced when an element is made to emit its characteristic light; this can be done by vapourising it, as, for example, when a piece of asbestos soaked in common salt solution is put in a bunsen flame, when it emits the well-known yellow light due to the element sodium. The spectrum of this light consists of two lines very close together in the yellow, called the D lines, no other visible light being emitted. A better source of sodium yellow light, or D light, is the lamp with two electrodes in the form of filaments, containing some sodium and a little argon. When the lamp is switched on, the current goes through each filament in series and the electrons emitted ionise the argon and render it conducting. The current is then made to pass through the gas from one filament to the other, the heat generated vapourising the sodium, which ultimately carries practically all the current. The sodium atoms are got into an excited state in this way, and in returning to their normal state they emit their characteristic D light. The gas hydrogen, and other gases such as argon, helium, and neon, can be made to give out their characteristic light by filling a tube with the gas at a suitably low pressure of a few millimetres of mercury and passing an electric current through the gas. All these gases give line spectra under these conditions, the hydrogen spectrum consisting of a line in the red, one in the blue green, one in the violet, and one in the extreme violet. Photographs of some spectra are shown in Plate VI, but the reader is advised to see as many as possible of them for himself. The important thing to realise is that here is the solution of Newton's problem supplied by Nature herself: the elements give sharply defined colours at a fixed point in the spectrum, which is just what Newton was seeking. Moreover, as we expect, the refractive index of hydrogen red light is always the same

for a given material. Further work on line spectra, which will be discussed in Chapter 16, has shown that they are produced by atoms.

Further investigation of the spectra of substances revealed a different kind of spectrum. If a tube is filled with nitrogen at a suitably low pressure and an electric current is passed through the tube, the spectrum of the light emitted is shown in Plate VI and has a fluted appearance. It is called a band spectrum, because it consists of a set of bands, which are diffuse at one side and end sharply at the other. Further investigation of band spectra with instruments of high resolving power (Art. 143) shows that each consists of a set of lines very close together, the separation of the lines decreasing as the sharp end of the band is approached. It is not possible to detect the individual lines with a prism spectrometer, as they are too close together. It will be shown later that such spectra are due to molecules, the nitrogen band spectrum being due to the nitrogen molecule, another good example being the spectrum of the compound cyanogen.

41. THE FRAUNHOFER LINES

Although Newton saw quite clearly how to produce a pure spectrum and did indeed make one, he seldom used it in his experiments on colour, preferring to use a circular aperture as his source, which gives a pure spectrum only at the top and bottom of the band of colour. It is natural that later investigators should work with a pure spectrum, as the view more clearly emerged that white light consists of a set of lights of different refractive index, and it was therefore desired to separate the different

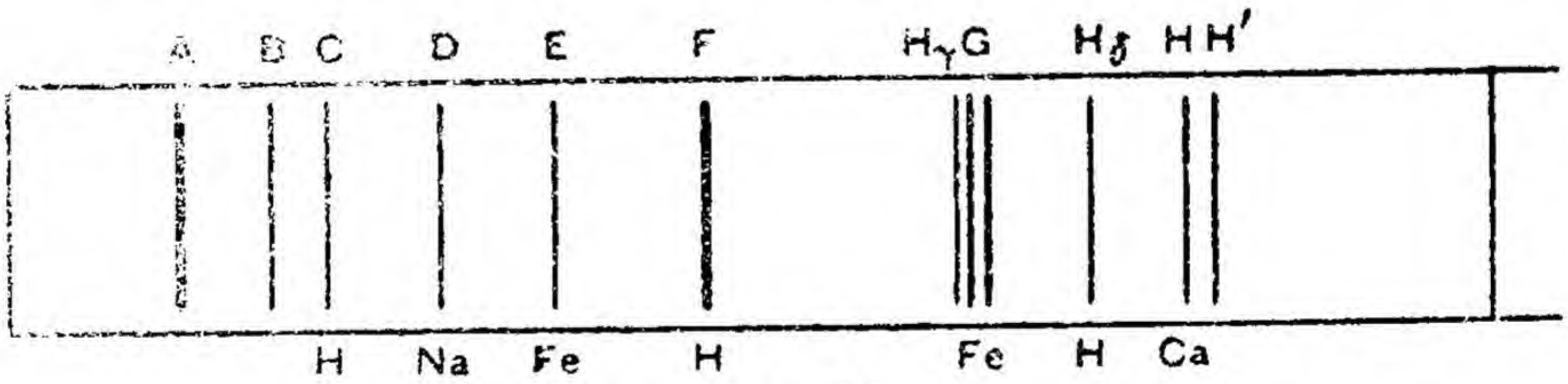


Fig. 61.

kinds of light as completely as possible. In producing such a spectrum using sunlight, Wollaston noticed in 1802 that it was crossed by a number of dark lines, but he offered no explanation of the phenomenon, which therefore went unnoticed for a time. In 1814 Fraunhofer, who was working in an optical firm in Munich and was particularly interested in the construction of good achromatic telescopic objectives, set up a pure spectrum using sunlight. This was one of a series of experiments designed to tackle Newton's old problem of fixing definite colours in the spectrum in order to get an accurate measure of the dispersive power of different materials, and to his surprise he saw that the spectrum, instead of being a continuous band of colour from red to violet, was crossed by a large number of dark lines. The appearance is represented in Fig. 61 and can readily be obtained in the laboratory by pointing the slit of a

spectrometer at the sun. Fraunhofer observed 576 of these lines and gave them letters after discovering that they always occurred in the same place in the spectrum. He also satisfied himself that they did not occur at the boundary between two colours, as was suggested at first, because the colour on the two sides of any particular line was always the same. No doubt the idea that the spectrum consists of a finite number of separate colours arose from Newton's attempt to separate the spectrum up in this way, but Fraunhofer's discovery finally disposed of this idea and proved that sunlight consists of a set of different kind of lights, whose refractive index shows an infinite variation between two given limits. Fraunhofer was not able to explain the significance of these dark lines, although he did show that the D line in the yellow coincides in position with the yellow line emitted by sodium. He did this by arranging for one-half of the slit of his spectroscope to be covered by an image of the sun, while the other was illuminated by sodium yellow light, so that the two spectra appeared one above the other. He observed the spectra through a telescope and saw that the bright yellow line in the sodium spectrum coincided exactly with the dark line which he had called D in the solar spectrum. This is why sodium yellow light is usually called D light. But Fraunhofer at once saw that his discovery had solved his difficulty, for he could use the dark lines as defining definite colours or places in the spectrum and so he could get an accurate measure of the dispersive power of different materials and improve the achromatism of his telescopic objectives. The dispersive power of a material is now defined as $\frac{n_F - n_C}{n_D - 1}$, where n_C, n_F, and n_D are the refractive indices of the material for the C, F, and D lines in the solar spectrum respectively. The D line is chosen as giving the mean refractive index of the material for the whole spectrum, and the C and F lines, in the red and blue-green respectively, are chosen because they cover the part of the spectrum to which the eye is most sensitive.

Before discussing in detail how achromatic lenses can be made, the physical significance of the Fraunhofer lines will be dealt with. It was soon discovered that not only was the D line identical in position with the yellow line in the sodium spectrum, but that the C, F, H_γ, and H_δ lines are also identical in position with the four lines in the emission spectrum of hydrogen, and that the H and K lines in the violet agree with two lines in the spectrum of calcium. This work was continued until the majority of the 576 lines discovered by Fraunhofer were identified as coinciding with emission lines in the spectra of sodium, hydrogen, calcium, iron, barium, magnesium, manganese, chromium, nickel, cobalt, aluminium, zinc, and copper. The explanation of this striking result was finally published by Kirchhoff in 1859, although it is in the notebook of a student attending Kelvin's lectures five years earlier, which can still be seen at Glasgow; it is said that Kelvin refused to publish the result as he

considered that Stokes really originated the idea, but Stokes insisted on being equally modest. Kirchhoff put a bunsen flame between the slit of a spectrometer and the sun and placed a piece of asbestos soaked in common salt in the flame. The D line in the solar spectrum was darker when the sodium flame was there than in its absence. He then replaced sunlight by a source of white light and observed the usual continuous spectrum; when the sodium flame was placed between the source of white light and the slit of the spectrometer, a dark line appeared in the same place as the D line in the solar spectrum. If the white light was removed, the dark line changed to a bright yellow line. It is evident that this dark line is due to the fact that sodium vapour absorbs just that light which it emits and no other; the line is not absolutely dark but only relative to the bright background produced by the white light, and, as soon as this background is removed, the sodium produces a bright yellow line. It follows that the line will only look dark if the sodium vapour is cooler than the source of white light, otherwise it will be emitting as much light by virtue of its own temperature as the source of white light is emitting. This view of the dark line in the continuous spectrum was verified by filling a tube with cool sodium vapour in hydrogen; when the tube was viewed through a source of sodium light it appeared dark, whereas through a source of white light it appeared bright. Dark lines corresponding to the emission lines in the spectrum of other elements can be produced in the same way and the resulting spectrum is called an absorption spectrum. The explanation of the Fraunhofer lines is now clear. The central portion of the sun is so hot that it gives out white light, which would produce a continuous spectrum. But this central core, the photosphere, is surrounded by an atmosphere consisting of vapours of various elements at a lower temperature than the photosphere; these elements absorb just those colours which they emit, and so the continuous spectrum is crossed by a set of dark lines identical in position with the emission lines of these elements. Kirchhoff supported his experimental results by the law named after him, which he derived from the second law of thermodynamics, that the ratio of the emissive power to the absorptive power of a substance for any given kind of light is the same for all substances. This means that, if a substance fails to emit certain colours, it will not absorb them at all; also it will absorb those colours which it emits. So, if a layer of cool sodium vapour is placed between a source of white light at a higher temperature and the slit of a spectrometer, it will absorb the D light, allowing the rest to go on. Of course, it follows from Prévost's theory of exchanges that the cool sodium vapour will be so excited by the absorption of D light from the incident light, until it emits as much as it is absorbing. But this emission occurs equally in all directions, and so the amount of D light restored to the original beam is less than taken from it when equilibrium is established, and so a dark line appears in the spectrum in the yellow. It should

be noticed that this line is not absolutely dark, as the vapour is sending some D light on to the slit of the spectrometer, but it is not sending the full amount emitted by the source. If the sodium vapour is at the same temperature as the source of white light, no dark line will be seen, since the sodium vapour is already emitting as much D light as the source of white light, and so it restores to the beam in the direction of the slit as much as it takes away. So the Fraunhofer lines prove that the sun is surrounded by an atmosphere cooler than the interior of the sun. Also the condition for seeing an absorption spectrum clearly is to have the absorbing substance as much cooler than the source of white light as possible.

The discovery that each element has its own characteristic line or band spectrum and that only a very small mass of the element is needed to produce its spectrum suggested the possibility of the discovery of new elements by means of their spectra. The method consisted in causing salts of known elements to emit their spectra, which were then carefully examined. If any lines were discovered which did not appear in the spectra of known elements, this suggested the presence of a new element in the salt, which could then be isolated chemically. It is interesting to notice that this sensitiveness of spectroscopic analysis obstructed its progress at the beginning. The first line spectrum to be discovered was the yellow line now known to be due to sodium, and it is estimated that only $\frac{1}{100{,}000}$ of a milligram of the element is necessary to produce this line in a spectrum. The result was that the D line appeared in nearly every spectrum, and this made it difficult to find the element which was responsible for it. The test was too sensitive! But this method of searching for new elements led to the discovery of caesium, rubidium, and thallium, all of which were afterwards isolated chemically. The method was finally applied with success to the solar spectrum itself by Lockyer. He observed certain dark lines, which could not be identified with those of any known element which had been produced in the laboratory, and he eventually came to the conclusion that they were due to a new element, which he called helium, which means the sun substance. Some years later the element was separated from the mineral cleveite and it is also found in certain natural waters and hot springs. The α-particles emitted by certain radio-active elements are helium nuclei. Helium is of great importance commercially as it is used to fill airships, being preferable to hydrogen on account of its much greater safety, its buoyancy being little inferior to that of hydrogen.

42. ACHROMATIC AND DIRECT VISION PRISMS

Newton's statement, that the dispersive power of all materials is the same, has been disproved by the discovery of line spectra, thus enabling definite colours to be employed for the measurement of the refractive

index and dispersive power of materials. No doubt also a wider range of materials was at the disposal of his successors than was available to Newton. This reversal of Newton's statement is no reflection on his genius; his conclusion was a sound one in the light of the materials and facilities at his disposal; both experimental fact and theory are continually being revised in the progress of science, the revision of facts being due to increased experimental skill and the overthrow of theories to the discovery of new facts, which cannot be embraced by the older theories. It is therefore possible to construct an achromatic lens, but, before proceeding to discuss how this can be done, the case of an achromatic prism will be treated.

An achromatic prism usually consists of a crown glass and flint glass prism in contact designed so that all colours are deviated through the same angle and so a beam of white light emerges from the prism as white light. If a beam of white light is sent on to a prism of crown glass, it produces a spectrum and let the deviations for a colour in the blue, one in the red, and for the D line be δ_1, δ_2, and δ respectively. Then $\delta_1-\delta_2$ is called the *dispersion* of the given prism for the two given colours, for it measures how much the two given colours are separated by the prism. This quantity depends both on the colours and also on the angle of the prism and on the angle at which the incident light strikes it. To put it in another way, it depends on the angle through which the beam as a whole has been deviated, and the greater this angle, the greater the dispersion of the two colours. If a quantity is to be formed which is a measure of the dispersive power of the material of the prism, a quantity characteristic of the material but independent of the dimensions of any particular prism, the dispersion of the two colours for a given deviation of the D line must be chosen. The unit deviation is selected, and so **the dispersive power of a material for two colours is the dispersion of those colours per unit deviation of the D line.** That is, the dispersive power, ω, is given by

$$\omega=\frac{\delta_1-\delta_2}{\delta} \qquad (23)$$

Unfortunately it is found that even this quantity varies with the angle of the prism, if it becomes large, but, if only prisms of small angle and rays incident at small angles are considered, the expression is constant. For, if α is the angle of the prism and n, n_1, and n_2 are the refractive indices of its material for the D line and the colours in the blue and red respectively, from equation (14)

$$\delta_1=(n_1-1)\alpha,\ \delta_2=(n_2-1)\alpha,\ \delta=(n-1)\alpha$$

the prism being in air.

$$\therefore\ \delta_1-\delta_2=(n_1-n_2)\alpha$$

$$\therefore\ \omega=\frac{n_1-n_2}{n-1} \qquad (24)$$

This quantity *is* constant for a given material and, since it is equal to the dispersion per unit deviation for the two colours produced by prisms of small angles, it is taken as defining the dispersive power of the material for the two given colours.

It is now required to find the angle α' of a prism of flint glass of dispersive power ω', which will form an achromatic combination with the above crown glass prism. This means that the two given colours are to be deviated through the same angle by the combination of prisms, but that some deviation of the D light must remain. Flint glass has a greater refractive index and dispersive power than crown glass. If the refractive index of flint glass for the three given colours is denoted by the same letters as before with a dash, then the deviations δ_1', δ_2', and δ' for blue, red, and the D line respectively are given by

$$\delta_1'=(n_1'-1)\alpha', \quad \delta_2'=(n_2'-1)\alpha', \quad \delta'=(n'-1)\alpha'.$$

For achromatism

$$\delta_1+\delta_1'=\delta_2+\delta_2'.$$

$$\therefore\ (n_1-1)\alpha+(n_1'-1)\alpha'=(n_2-1)\alpha+(n_2'-1)\alpha'$$

$$\therefore\ (n_1-n_2)\alpha+(n_1'-n_2')\alpha'=0 \qquad \ldots\ldots\ldots \qquad (25)$$

from which a unique value of α' is determined, the negative value indicating that the second prism must be placed with its refracting edge pointing in the opposite direction to that of the first prism, so that it deviates the light in the opposite sense. To show that this destruction of dispersion does not destroy deviation, the deviation of the D light in each prism is given by

$$\delta=(n-1)\alpha \text{ and } \delta'=(n'-1)\alpha'$$

Substituting these values of α and α' in equation (25)

$$\frac{(n_1-n_2)}{(n-1)}\delta+\frac{(n_1'-n_2')}{(n'-1)}\delta'=0$$

$$\therefore\ \omega\delta+\omega'\delta'=0$$

That is, if $\omega=\omega'$, $\delta+\delta'=0$ and so in destroying dispersion, deviation is also destroyed, as Newton realised. But if $\omega\neq\omega'$, $\delta+\delta'\neq 0$, and so some deviation remains. Since $(n_1'-n_2')$ is greater than (n_1-n_2) the angle of the flint glass prism is less than that of the crown glass prism. Fig. 62 shows that, if a single ray of light is sent through the achromatic prism, the emergent beam will consist of the colours of the spectrum all travelling along parallel rays, but, if a beam of finite width is used, these colours will overlap to produce white light except at the edges. This treatment is only true for prisms of small angle with light falling on at nearly normal incidence. If a prism of

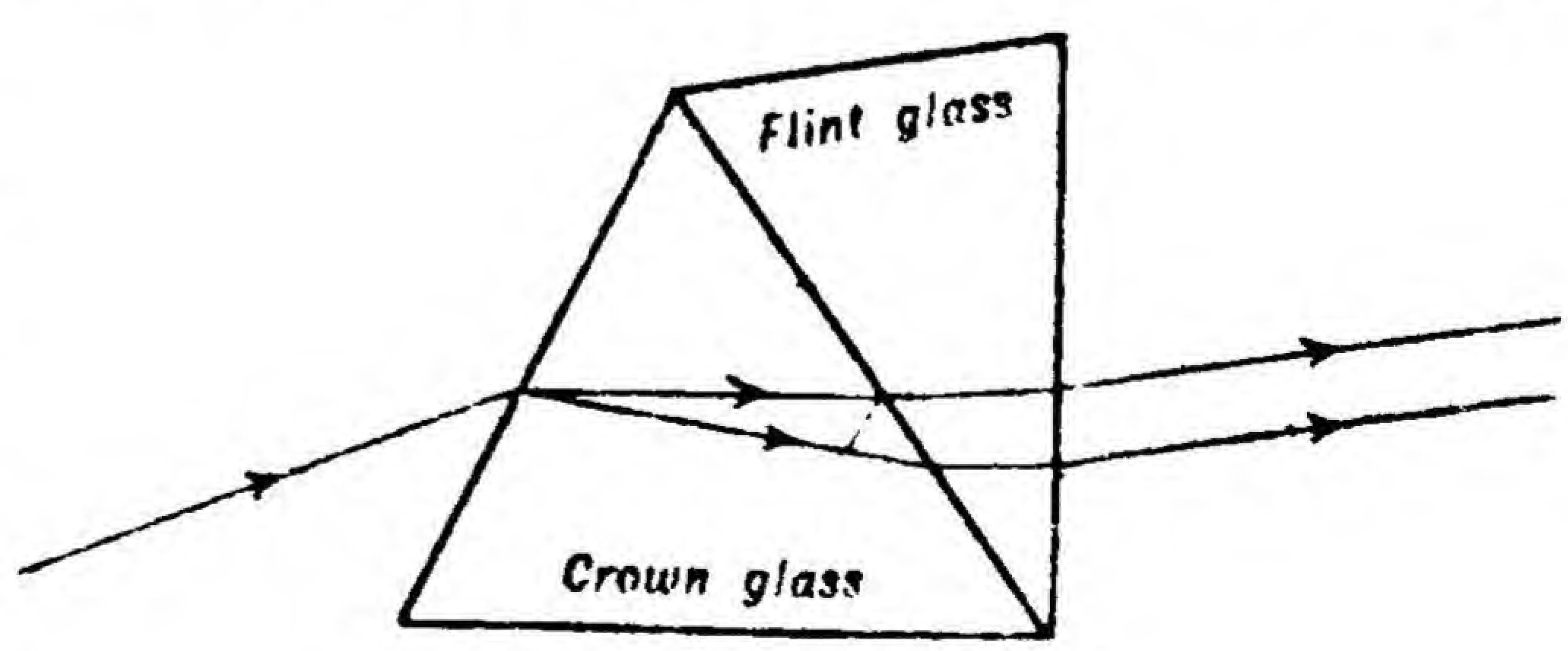


Fig. 62.

flint glass is to be combined with a crown glass prism of finite angle to produce an achromatic combination, the deviations must be worked out for a given angle of incidence on the crown glass prism and the combination will only be achromatic round and about such angles.

It must be emphasised that even for prisms of small angle, the combination is only achromatic for the two colours considered. Other colours will emerge from the combination not quite parallel to these. But what two colours are to be chosen to give the best results ? This is settled by trying various pairs of colours and seeing which gives the best result in practice, and it is found that a combination achromatic for the C and F lines is the best for visual work, and so the **dispersive power ω of a material is defined as the dispersion of the C and F lines per unit deviation of the D line**, or

$$\omega=\frac{n_F-n_C}{n-1}$$

where n_F and n_C are the refractive indices of the material for the F and C lines respectively. Presumably these two lines give the best results, because they cover the region of the spectrum, to which the eye is most sensitive.

Achromatic prisms are not of any use in practice, although the fact that they can be made shows that it is possible to make achromatic lenses, since a lens can be regarded as a set of prisms of small angle (Art. 17). A more useful combination is one which produces dispersion without deviation and is called a **direct vision prism**; it is used in a direct vision spectroscope, which serves for the quick qualitative examination of spectra. If a crown glass prism has an angle α, it is required to find the angle α' of the flint glass prism, which will combine with it so as to produce no deviation of sodium yellow light. Using the same notation for refractive indices as before,

$$(n-1)\alpha+(n'-1)\alpha'=0$$

$$\therefore\ \alpha'=-\left(\frac{n-1}{n'-1}\right)\alpha \quad . \quad . \quad . \quad . \quad . \quad (26)$$

The dispersion of the F and C lines is given by

$$\begin{aligned}\delta_F-\delta_C&=(n_F-1)\alpha+(n'_F-1)\alpha'-(n_C-1)\alpha-(n'_C-1)\alpha'\\&=(n_F-n_C)\alpha+(n'_F-n'_C)\alpha'\end{aligned}$$

Substituting the value of α' from equation (26) in this equation

$$\begin{aligned}\delta_F-\delta_C&=(n_F-n_C)\alpha-\frac{(n'_F-n'_C)(n-1)}{(n'-1)}\alpha\\&=\left(\frac{n_F-n_C}{n-1}-\frac{n'_F-n'_C}{n'-1}\right)(n-1)\alpha\\&=(\omega-\omega')(n-1)\alpha\end{aligned}$$

PLATE I

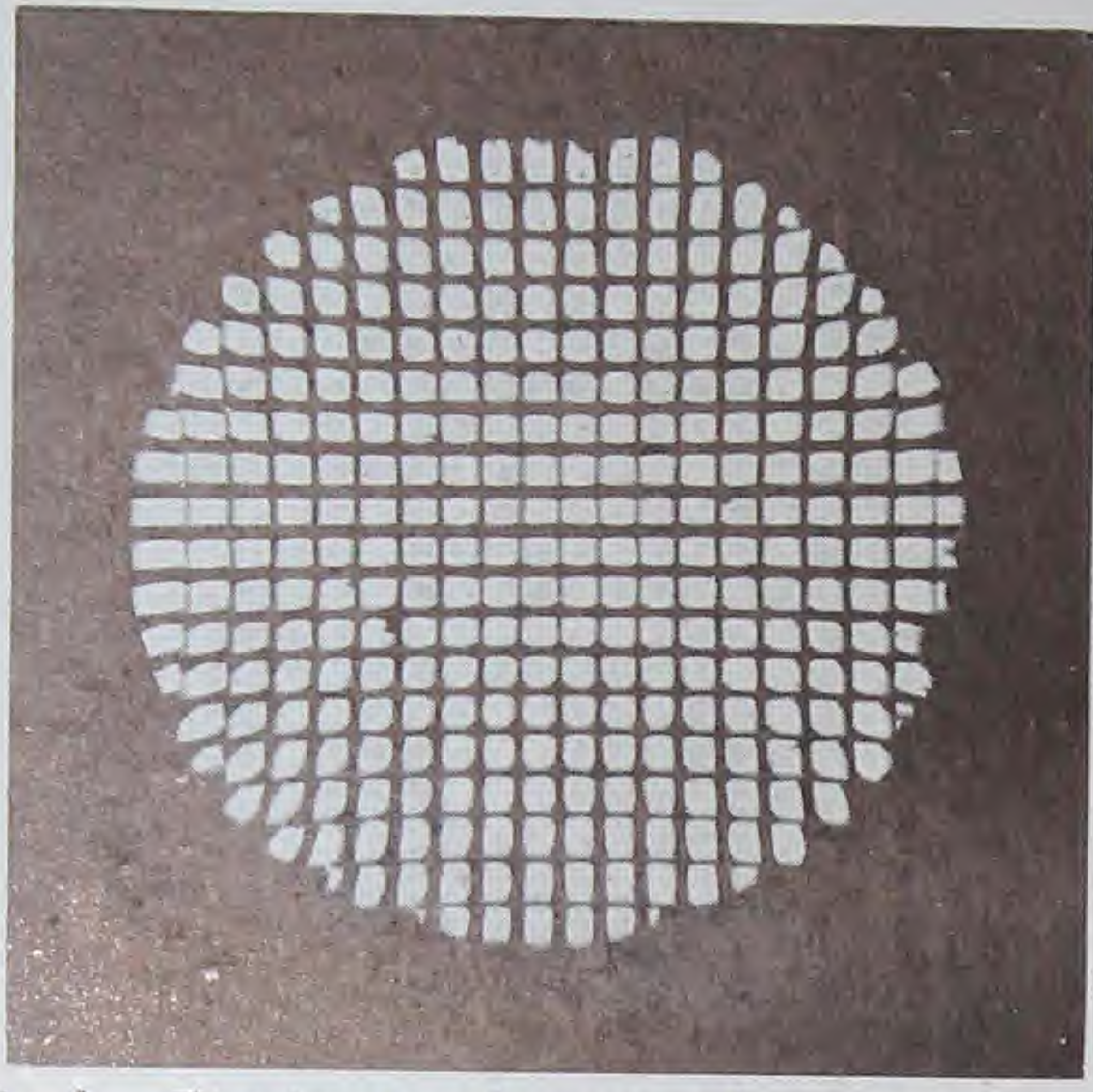


Fig. 1. The image of a gauze formed by an ordinary lens.

(*J. W. Cottingham*)

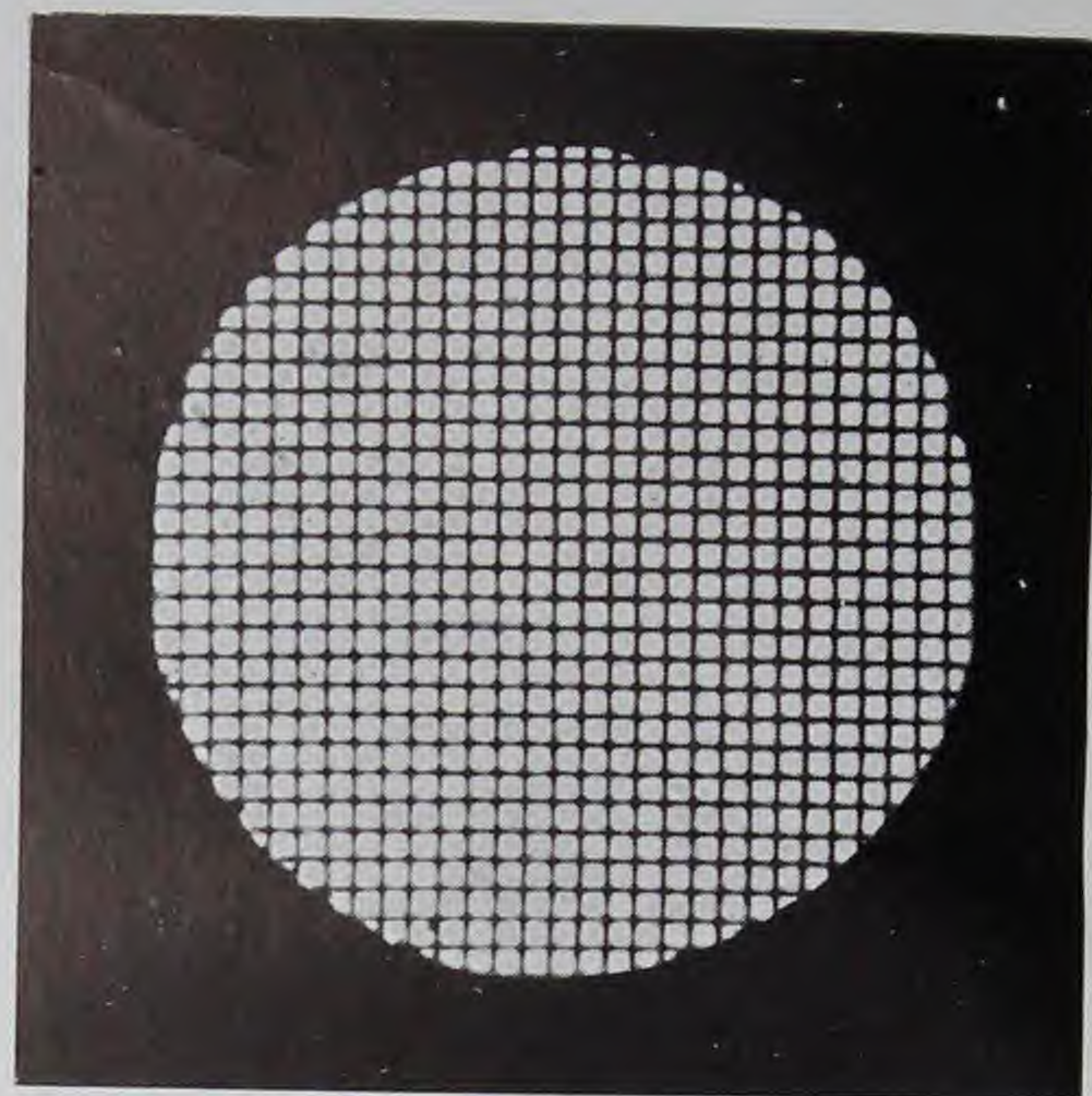


Fig. 2. The image of a gauze formed by an achromatic lens showing improved definition.

(*J. W. Cottingham*)

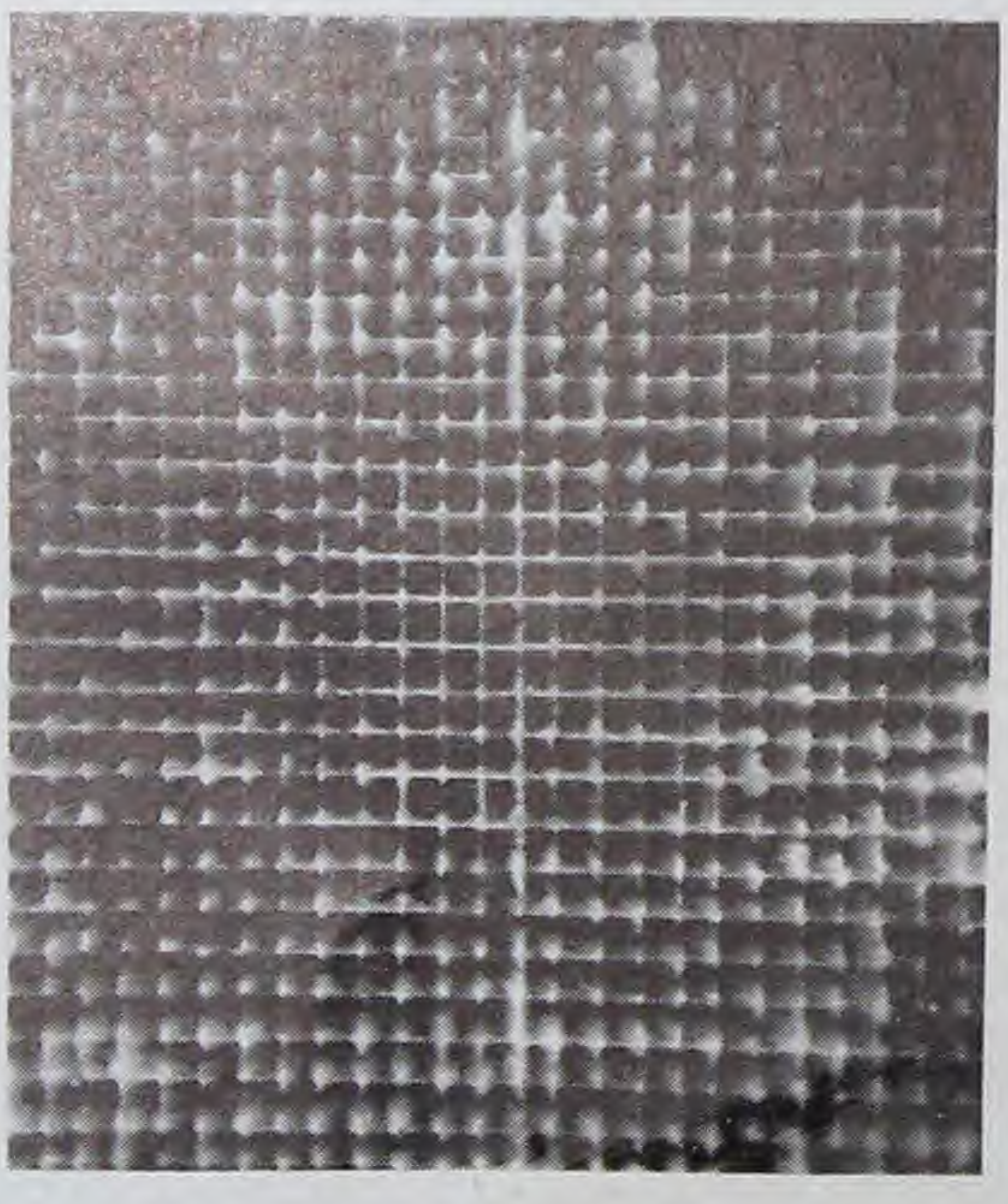


Fig. 3. The image of a gauze formed by using the whole of the aperture of a lens.

(*J. W. Mitchell*)

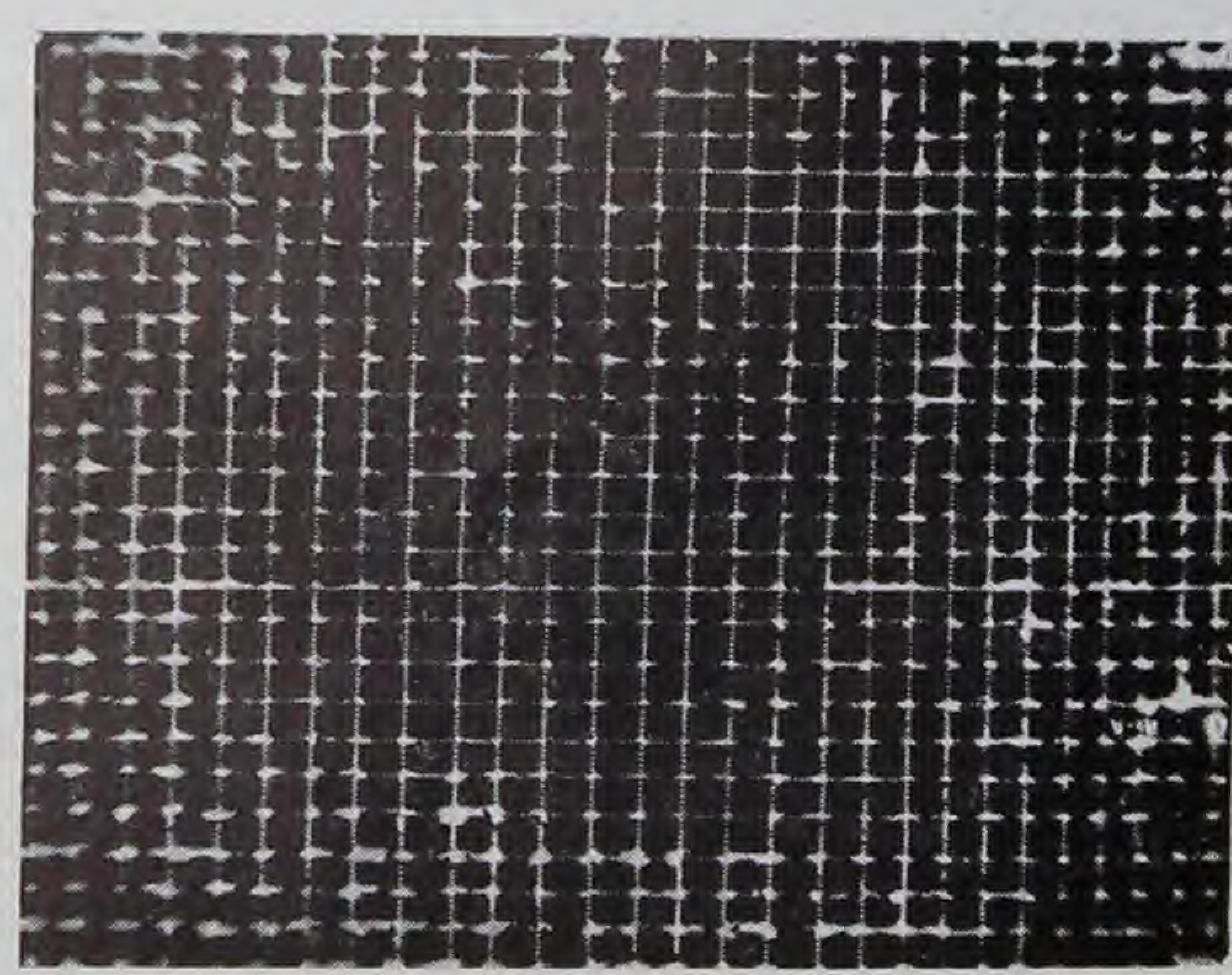


Fig. 4. The image of the same gauze formed by the same lens stopped down to f.11, showing the improved definition.

(*J. W. Mitchell*)

which is not zero if $\omega \neq \omega'$. If prisms of finite angle are to be used, then the angle of the flint glass prisms must be worked out for the particular angle of incidence at which the light enters the first prism.

43. ACHROMATIC LENSES

✓If a lens of given focal length is to be constructed so as to be achromatic for r colours, it must consist of r lenses in contact made of materials of different dispersive power. To consider the case of the H, F, and C line in the violet, green, and red, respectively, let F be the focal length of the required achromatic lens, n be the refractive index of the material of the first lens for the D line, f be the corresponding focal length, and r and s be the radii of curvature of its surfaces. Let the same letters with one and two dashes represent the corresponding quantities for the second and third lens respectively. Then

$$\frac{1}{f}=(n-1)\left(\frac{1}{r}-\frac{1}{s}\right)$$

$$\therefore \quad -\frac{df}{f^2}=dn\left(\frac{1}{r}-\frac{1}{s}\right)$$

where dn is the difference in refractive index between the H and F lines.

$$\therefore \quad -\frac{df}{f^2}=\frac{dn}{n-1}\cdot\frac{1}{f}$$

$$\therefore \quad -\frac{df}{f^2}=\frac{{}_{\mathrm{H}}\omega_{\mathrm{F}}}{f}$$

where ${}_{\mathrm{H}}\omega_{\mathrm{F}}$ is the dispersive power of the material for the H and F lines.

Again
$$-\frac{\delta f}{f^2}=\delta n\left(\frac{1}{r}-\frac{1}{s}\right)$$

where δn is the difference in refractive index between the F and C lines.

TABLE 5

	n_{H}	n_{F}	n	n_{C}	${}_{\mathrm{H}}\omega_{\mathrm{F}}$	${}_{\mathrm{F}}\omega_{\mathrm{C}}$	$\frac{{}_{\mathrm{H}}\omega_{\mathrm{F}}}{{}_{\mathrm{F}}\omega_{\mathrm{C}}}$
Crown glass	1·5330	1·5230	1·5170	1·5145	0·0194	0·0164	1·18
Flint glass	1·6883	1·6637	1·6499	1·6434	0·0378	0·0297	1·27

$$\therefore \quad -\frac{\delta f}{f^2}=\frac{{}_{\mathrm{F}}\omega_{\mathrm{C}}}{f}$$

where ${}_{\mathrm{F}}\omega_{\mathrm{C}}$ is the dispersive power of the material for the F and C lines. Now

$$\frac{1}{\mathrm{F}}=\frac{1}{f}+\frac{1}{f'}+\frac{1}{f''} \quad \ldots \ldots \quad (27)$$

For achromatism for the three colours, $dF=\delta F=0$.

Now
$$-\frac{dF}{F^2}=-\frac{df}{f^2}-\frac{df'}{f'^2}-\frac{df''}{f''^2}$$

and
$$-\frac{\delta F}{F^2}=-\frac{\delta f}{f^2}-\frac{\delta f'}{f'^2}-\frac{\delta f''}{f''^2}$$

Therefore, for achromatism

$$\frac{df}{f^2}+\frac{df'}{f'^2}+\frac{df''}{f''^2}=0$$

$$\text{and } \frac{\delta f}{f^2}+\frac{\delta f'}{f'^2}+\frac{\delta f''}{f''^2}=0$$

$$\therefore \frac{{}_{H}\omega_{F}}{f}+\frac{{}_{H}\omega'_{F}}{f'}+\frac{{}_{H}\omega''_{F}}{f''}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

and
$$\frac{{}_{F}\omega_{C}}{f}+\frac{{}_{F}\omega'_{C}}{f'}+\frac{{}_{F}\omega''_{C}}{f''}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

Given the three materials and the required focal length of the achromatic lens, f, f', and f'' are determined from equations (27), (28), and (29). Two lenses will not be sufficient, since that would mean that there were three equations to find two unknowns, which signifies that two of the equations must be equivalent to one another. This is not the case, since equations (28) and (29) are not equivalent, as $\frac{{}_{H}\omega_{F}}{{}_{F}\omega_{C}} \neq \frac{{}_{H}\omega'_{F}}{{}_{F}\omega'_{C}}$, which can be seen from Table 5 for a typical crown glass and flint glass. Thus, if the combination is made achromatic for violet and green, it is not corrected for green and red. Telescopic objectives which are to be used for photography are made achromatic for the C, F, and H_δ lines, the three hydrogen lines in the red, blue-green, and violet respectively. This covers the range of the spectrum to which the ordinary photographic plate is sensitive.

For visual work it is sufficient to make the lens achromatic for the C and F lines, the eye being very insensitive to the violet. It is at once apparent that equations (27), (28), and (29) are replaced by

$$\frac{1}{F}=\frac{1}{f}+\frac{1}{f'} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

and
$$\frac{\omega}{f}+\frac{\omega'}{f'}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

and only two lenses are necessary now. These equations determine f and f' given F, ω, and ω'. It is usual to have a crown glass convex lens, when it is clear that the flint glass component must be a concave lens. A practical optician specifies a lens by its material and the radii of curvature

of its faces rather than by its focal length. To do this we have the following additional equations :

$$\frac{1}{f}=(n-1)\left(\frac{1}{r}-\frac{1}{s}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

$$\frac{1}{f'}=(n'-1)\left(\frac{1}{s}-\frac{1}{s'}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

It will be noticed that the radii of curvature of the faces of the individual lenses which are to be in contact have been made equal, so that they can be cemented together. Since f and f' are determined by equations (30) and (31), there are two equations to determine the three unknowns r, s, and s'. So there is the possibility of imposing yet another condition on the system, which can be used to make the spherical aberration a minimum and then all the four unknowns are uniquely determined. The focal length of the combination is given by equation (30) and is not infinity if $\omega \neq \omega'$, so that in destroying the dispersion we have not eliminated the deviation. We now see that it is possible to extend the design of a projection lens considered in Art. 16 to a lens combination free from chromatic aberration.

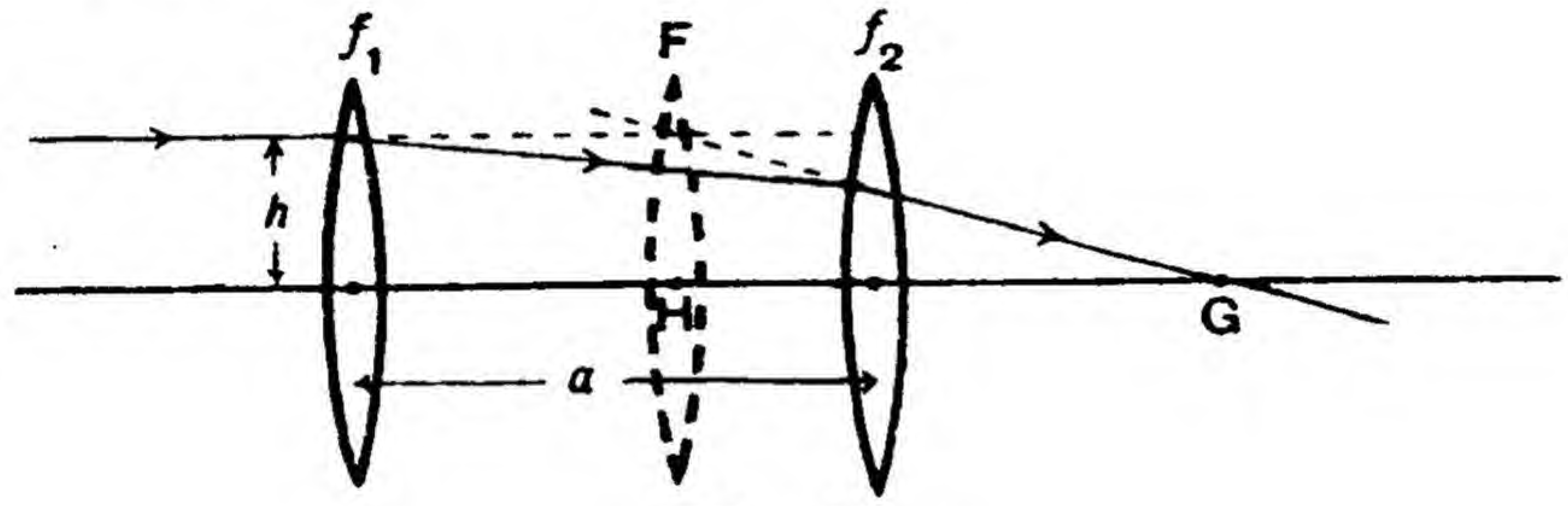


Fig. 63.

It is interesting to notice that it is possible to make an achromatic combination with two lenses of the same material, if they are separated by a finite distance. To see how this is done we must first calculate the focal length of such a combination. If the two lenses have a focal length f_1, f_2 respectively, and are a distance a apart (Fig. 63), a ray parallel to their common axis at a distance h from it is deviated through an angle $\frac{h}{f_1}$ by the first lens and strikes the second lens at a distance $\left(h-\frac{ah}{f_1}\right)$ from the axis. It has been proved in Art. 17 that the deviation in a ray produced by a lens is independent of the angle at which it strikes it, provided it is small, and so the second lens deviates the ray through a further angle $\dfrac{\left(h-\frac{ah}{f_1}\right)}{f_2}$. The ray emerges from the second lens and crosses the axis at an angle $\frac{h}{f_1}+\frac{1}{f_2}\left(h-\frac{ah}{f_1}\right)$ or $\frac{h}{f_1}+\frac{h}{f_2}-\frac{ah}{f_1 f_2}$. If a single lens is to be used,

which is equivalent to this combination as far as a ray parallel to the axis is concerned, it must be placed at the point H and have a focal length F equal to GH.

$$\therefore \frac{h}{F}=\frac{h}{f_1}+\frac{h}{f_2}-\frac{ah}{f_1 f_2}$$

$$\therefore \frac{1}{F}=\frac{1}{f_1}+\frac{1}{f_2}-\frac{a}{f_1 f_2} \quad \ldots \ldots \ldots \quad (34)$$

This expression is true for converging and diverging lenses, provided that the appropriate sign is given to the focal length of each lens. It follows from Art. 17 that it is true for rays making any small angle with the axis.

Now
$$-\frac{dF}{F^2}=-\frac{df_1}{f_1^2}-\frac{df_2}{f_2^2}+a\left(\frac{df_1}{f_1^2 f_2}+\frac{df_2}{f_1 f_2^2}\right)$$

$$\therefore -\frac{dF}{F^2}=\frac{\omega}{f_1}+\frac{\omega}{f_2}-\frac{2a\omega}{f_1 f_2},$$

where ω is the dispersive power of the material of the lenses.
For achromatism $dF=0$.

$$\therefore \frac{\omega}{f_1}+\frac{\omega}{f_2}=\frac{2a\omega}{f_1 f_2}$$

$$\therefore a=\frac{f_1+f_2}{2} \quad \ldots \ldots \ldots \quad (35)$$

The problem of making an achromatic telescopic objective in order to improve the sharpness of the images formed by telescopes has at last been solved, and it is interesting to recall the stages by which this has been done, and to see what other fresh ground has been broken and what further lines of advance suggest themselves. It was this trouble with the blurred and coloured images formed by telescopes which led Newton to his experiments on the passage of sunlight through a prism. They resulted in the conception that white light consists of a set of independent colours, the colours of the spectrum, and to the association of colour with refractive index. This second result is the more important of the two; it marks the replacement of the physiological sensation of colour by the physical quantity refractive index. This is always the first stage in introducing scientific method into any new realm, and at this stage the facts can be changed from subjective to objective and the physical quantity usually has a much wider range than the physiological sensation which it has replaced. For example, the physical quantity temperature measures a much wider range of degrees of hotness than can be detected by the sense of touch and thereby greatly extends the scope of scientific investigation in the phenomena of heat.

Newton's work on the spectrum supplied the explanation of the blurred

images of telescopes, but the limited range of materials at his disposal led him to believe that it was impossible to eliminate the defect. Fraunhofer returned to the same problem just over a hundred years later, and, while attempting to solve it, he discovered the dark lines in the solar spectrum and saw that they were just what was needed for the accurate measurement of the dispersive power of materials, and by their means he produced achromatic telescopic objectives. He himself did not discover the significance of the lines, but his discovery led to the finding of the corresponding emission lines and to the idea of each element having its own spectrum. This led to the discovery of new elements and finally to the explanation of the dark lines as due to an atmosphere of cool vapours round the sun, the elements in which absorbed just those colours which they emit. So it became possible to discover the elements in the atmosphere of both the sun and the stars. The reader is reminded that all this work began in an attempt to solve a practical problem, how to improve telescopes. And its continuation was inspired by the same motive, Fraunhofer being employed in a glass works at Munich where telescopes were made. This is a case where the demands of industry stimulated workers both in pure and applied science, and the solution of the problem led to discoveries and work of the greatest value and interest in pure knowledge. In the relations between science and industry it is a mistake to think that the debt is all on the side of industry. While it is true that the knowledge discovered by men interested only in knowledge for its own sake and indifferent to practical application has revolutionised industry, and that scientific method has improved existing processes, it is also true that pure science, in its turn, owes the opening up of new fields of pure knowledge to the problems put to it by industry. Indeed, it is essential that science should remain in close contact with the practical problems of the day if it is not to become sterile and dead.

Finally, two new problems suggest themselves. While the eye is used, the spectrum has definite limits ; but if colour is replaced by refractive index, no such limitation suggests itself. Why should there not be rays less refrangible than the red and more refrangible than the violet ? They will not affect the eye, it is true, but that does not affect their physical reality. A search must be made for such rays. Secondly, the physical quantity specifying a definite colour, such as the D line, is its refractive index. But this number varies according to the sort of glass or material we use ; it is 1·51 in a typical crown glass, 1·6 to 1·9 in flint glass, 1·632 in carbon disulphide. We cannot be satisfied to specify a definite colour with a number which changes in this way. At present it is the best we can do, but in all our investigations of light we must be looking for something better, for some way of measuring colour by a number which is invariable.

EXAMPLES ON CHAPTER V

(Use the refractive indices given in Table 5 unless they are actually given in the question itself)

1. Describe and discuss any phenomena known to you in which the direction or nature of a beam of light is affected by passage through water. (*Oxford Schol.*)

2. What adjustments are necessary before using a prism spectroscope? Give a concise explanation of the purpose of each of these adjustments. (*Oxford Schol.*)

3. Describe and explain how you would produce in the laboratory a parallel beam of monochromatic light and how you would demonstrate that it had these properties. (*Oxford. Schol.*)

4. Give a brief account of the more important uses of the spectrometer. (*Camb. Schol.*)

5. Describe the optical system employed in the prism spectrometer.

Give some account of the types of spectra emitted by different classes of luminous bodies. (*Camb. Schol.*)

6. Describe the adjustment and use of a prism spectrometer to measure refractive indices, and explain how some of the principal sources of error may be eliminated. (*Camb. Schol.*)

7. Explain how to produce a pure spectrum. Give an explanation of the dark lines in the solar spectrum stating what principle they illustrate. (*Camb. Schol.*)

8. Describe the optical system of the spectrometer, and the adjustments you would make in setting up the instrument to measure the refractive index of a prism.

Find the relation between the angle of the prism, the minimum deviation produced, and the refractive index of the material. (*O. and C.*)

9. Describe the spectrometer, explaining clearly the functions of its constituent parts. (*O. and C.*)

10. What is meant by the minimum deviation of a ray of light refracted through a prism? Explain clearly how the angle of minimum deviation for a given prism depends upon the colour of the light.

A ray of light passes through a prism whose refracting angle is 10° at approximately minimum deviation. If the refractive index of the glass is 1·514 for red light and 1·532 for blue light, determine the angular dispersion produced by the prism. Prove any formula you employ. (*London.*)

11. Describe how you would use a spectrometer to determine the refractive index of glass. Draw a diagram to show the optical arrangement of the instrument, indicating the paths of a pencil of rays from a source of monochromatic light.

PQR is a right-angled glass prism, the angles at P and R being 60° and 30° respectively. The face QR is silvered and a ray of light directed on to the face PR at an angle of incidence of 50° retraces its path after reflection from QR. What is the refractive index of the glass? What would be the angle of minimum deviation for a 60° prism made of the same glass? (*N.U.J.B.*)

12. A ray of light is incident at nearly normal incidence upon a thin glass prism. Show that its deviation is given by $(\mu-1)i$, where i is the angle of the prism and μ is its refractive index. For such a prism in air the deviation is found to be 4° towards the base of the prism. The prism is put into a rectangular glass tank filled (a) with water, (b) with carbon bisulphide. In (a) the deviation is 1·6° towards the base and in (b) 0·8° away from the base of the prism. Taking the refractive index of water to be 1·3, calculate the angle of the prism, the refractive index of the prism and the refractive index of carbon bisulphide. (*Camb. Schol.*)

13. Say how you would determine the refractive index of a glass prism, stating the chief sources of error.

Light from an illuminated slit placed in the focal plane of a thin lens of focal length 20 cm. falls on one side of a prism (vertical angle 30°) so that the refracted ray meets the second face, which is silvered, at approximately a right angle. For

violet light, the reflected image coincides with the slit, for red light it falls 5·58 mm. to the side of the slit. If the refractive index for violet light is 1·532, what is the refractive index for red light ? (*Oxford Schol.*)

14. A parallel beam of white light is incident on a 60° crown glass prism, which is set at minimum deviation for the D line. Find the angle between the beams of C and F light when they emerge from the prism. Through what angle must the prism be turned to make the F light go through it at minimum deviation ?

15. A parallel beam of white light is incident on a 60° flint glass prism, which is set at minimum deviation for the F line. The specrum is cast on a screen by a lens of 50·0 cm. focal length. Find the length of the spectrum between the C and F lines, the C and D lines, and the D and F lines.

16. A flint glass prism is to be combined with a crown glass prism of 3° so as to produce no deviation for the D line when passing through the combination at about normal incidence. Find the angle of the flint glass prism and the angle between C and F light when emerging from the combination.

17. What is meant by deviation and dispersion ?

"In a direct vision spectroscope we have dispersion without deviation, and in a lens corrected for chromatic aberration we have deviation without dispersion."

Explain these statements and describe the instruments mentioned. (*Oxford Schol.*)

18. State clearly how dispersion of light may be obtained without deviation and deviation without dispersion. What practical use is made of this knowledge ? (*Camb. Schol.*)

19. Explain the dispersion produced by a simple lens, and show how the defect may be corrected.

Why is such a correction unnecessary in the case of a simple convex lens used as a magnifiying glass held close to the eye ? (*O. and C.*)

20. Deduce an expression connecting the distances of an object and its image from a lens with the focal length of the lens.

It is desired to make a converging achromatic lens, of mean focal length 30 cm., by using two lenses of materials A and B. If the dispersive powers of A and B are in the ratio 1 to 2, find the focal length of each lens. (*O. and C.*)

21. Describe the optical principles of the construction of either a direct vision spectroscope or a simple achromatic photographic lens combination. Briefly describe an experiment you have seen which illustrates the optical principle involved in the construction of the instrument you choose. (*London.*)

22. What is meant by the dispersive power of a glass ? Show how lenses can be combined to give a combination which has the same focal length for two different colours. (*London Inter.*)

23. Parallel white light is incident on a 60° flint glass prism set at minimum deviation for the D line. The angle between the parallel beams of D light and F light on emerging from the prism is 2·3°. If the refractive index of the D light is 1·632, find that of the F light.

24. A 60° crown glass prism is to be combined with a flint glass prism so that there is no deviation for the D line, each prism being traversed at minimum deviation. Find the angle of the flint glass prism and the angle between the C and D lines on emerging from the combination.

25. Find the focal length for the D line of a flint glass lens which will form an achromatic combination for the C and F lines with a crown glass converging lens of focal length 20 cm. for the D line. Find also the focal length of the crown glass converging lens for the C and F lines and the focal length of the combination for the C, D, and F lines. Draw a graph of focal length against refractive index for each lens on the same axes to illustrate the degree of achromatisation.

26. A lens of focal length 6·00 in. for the D line is required for a camera and it is to be achromatic for the C and F lines. Calculate the focal lengths of the crown

and flint glass components; find also the radii of curvature of their surfaces if the two lenses are cemented together and the other pair of surfaces have also the same radius of curvature as one another.

27. The episcopic lens of an epidiascope is required to cast on a screen an image 8 ft. by 8 ft. of a plate, 10 in. by 10 in., which is 12 ft. from the screen. The lens is to be achromatic for the C and F lines; calculate the focal length of the crown and flint glass components and also the radii of curvature of their surfaces, if each lens has two surfaces whose radii are numerically equal in magnitude but opposite in sign.

28. It is required to produce a portrait lens of focal length 12·00 in., which is to be achromatic for the C, F, and H lines. Calculate the focal length of the components A, B, and C, whose dispersive powers for the H and F, and F and C lines respectively are, A : 0·0194, 0·0164; B : 0·0290, 0·0238; C : 0·0378, 0·0297.

29. Two thin coaxial lenses of focal lengths f_1 and f_2 are mounted at a distance d apart. Find the focal length of the equivalent thin lens.

Show how it is possible to construct an achromatic combination from two thin lenses of the *same* material. (*Camb. Schol.*)

30. Assuming the formula

$$\frac{1}{F}=\frac{1}{f_1}+\frac{1}{f_2}+\frac{d}{f_1 f_2}$$

show that two lenses of equal focal length and made of the same glass, can be arranged to give an achromatic system.

Draw, to scale, a diagram showing the paths of two rays of light of different wave-lengths which proceed from the same object. Take the focal lengths of the lenses to be 5 cm. for the one wave-length and 4·5 cm. for the other.

Discuss your result, explaining what you mean by an achromatic system. (*Camb. Schol.*)

Chapter VI

THE DEFECTS OF THE IMAGE FORMED BY A SINGLE LENS

44. SPHERICAL ABERRATION

We have seen that the theorems deduced for mirrors and thin lenses are only true for paraxial conditions, which implies that the mirror or lens must be of small aperture and the object must be small and close to the axis. We have also seen in Art. 26 that, if rays making a finite angle of incidence at the surface of a spherical mirror of large aperture are considered, the focal length of marginal rays is numerically less than that of paraxial rays and again that a pencil incident at a large angle on a mirror of small aperture gives rise to an astigmatic reflected pencil. Both of these cases will arise in practice for lenses as well as mirrors ; if a photographer wishes to take a picture of some moving subject on a dull day, he cannot admit the necessary amount of light to his plate by increasing the exposure, so he is compelled to enlarge the aperture of the lens and to admit rays, which fall on the lens at such large angles that paraxial conditions are not satisfied, even if the object be small and close to the lens. Again, if a photographer has to take a landscape, even on a bright day, the size of the picture may be such that, even though his lens is stopped down so that rays from the centre of the landscape are paraxial, those from its edges do not satisfy this condition. It is therefore essential to extend our analysis to include these cases, in order to see what type of image is to be expected from lenses and mirrors and to see how lenses of high quality for photographic and other purposes may be produced. We shall not consider the case of mirrors further, since they are of less practical importance than lenses and an exact mathematical treatment is tedious and yields little more insight into the nature of the image formed than has already been obtained. So we turn to the case of a point object on the axis of a lens of large aperture, which must naturally be approached through refraction at one spherical surface. We shall prove merely from *Snell's law and the spherical nature of the surfaces* that the focal length of marginal rays is numerically less than that of paraxial rays, whatever the type of lens. *While it is not necessary for the reader to learn this proof*, he is advised to go through it and to satisfy himself of its validity.

We have already seen (Art. 9) that, if a point source of light A on the

axis AOC of a spherical refracting surface sends a ray of light AD to the surface incident at angle i_1, it is refracted at an angle i_2 to meet the axis at B so that

$$\frac{n_2}{q}-\frac{n_1}{p}=\frac{n_2 \cos i_2-n_1 \cos i_1}{r} \quad . \quad . \quad . \quad . \quad . \quad (36)$$

where p, q, and r are equal to $-$AD, $+$BD, and $+$OC, the radius of curvature of the surface respectively (Fig. 64). This relation is true whatever

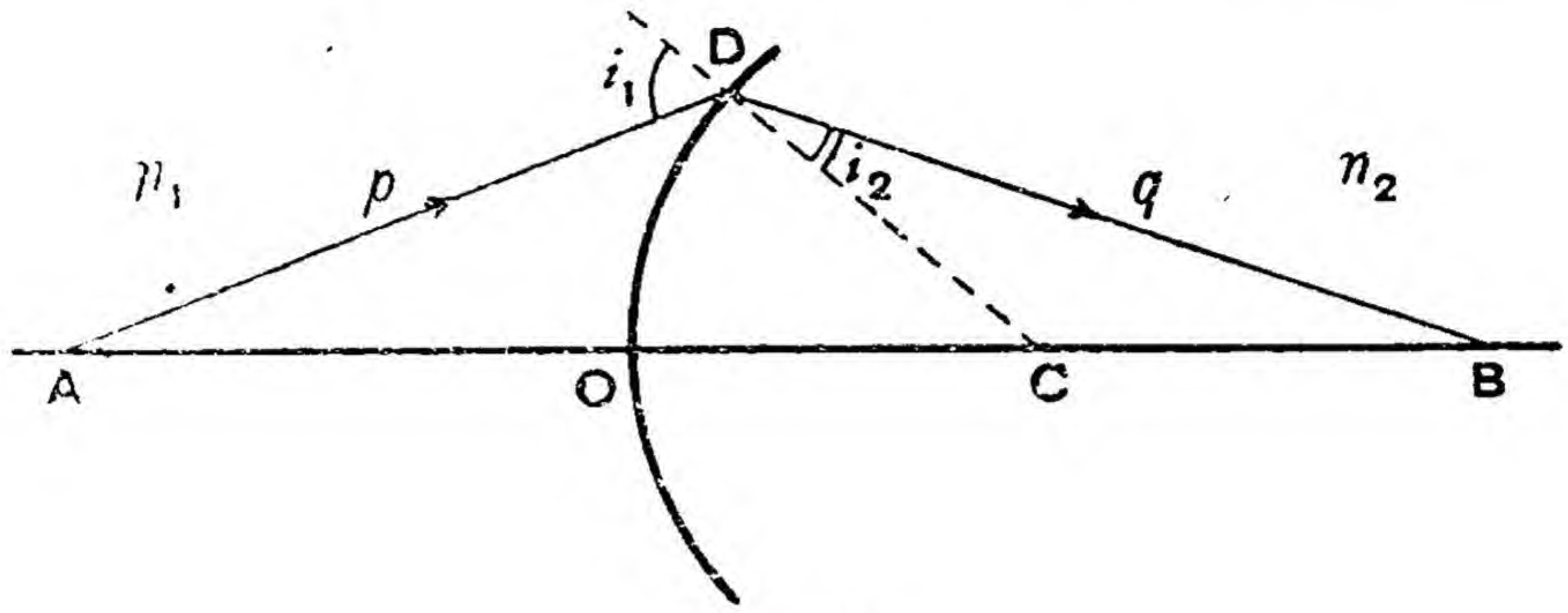


Fig. 64.

the size of the angles i_1 and i_2. Let us consider the simple case in which the point A is at infinity to the left of O on the axis AOC (Fig. 65), so that it sends a beam of rays parallel to the axis on to the refracting surface. Let

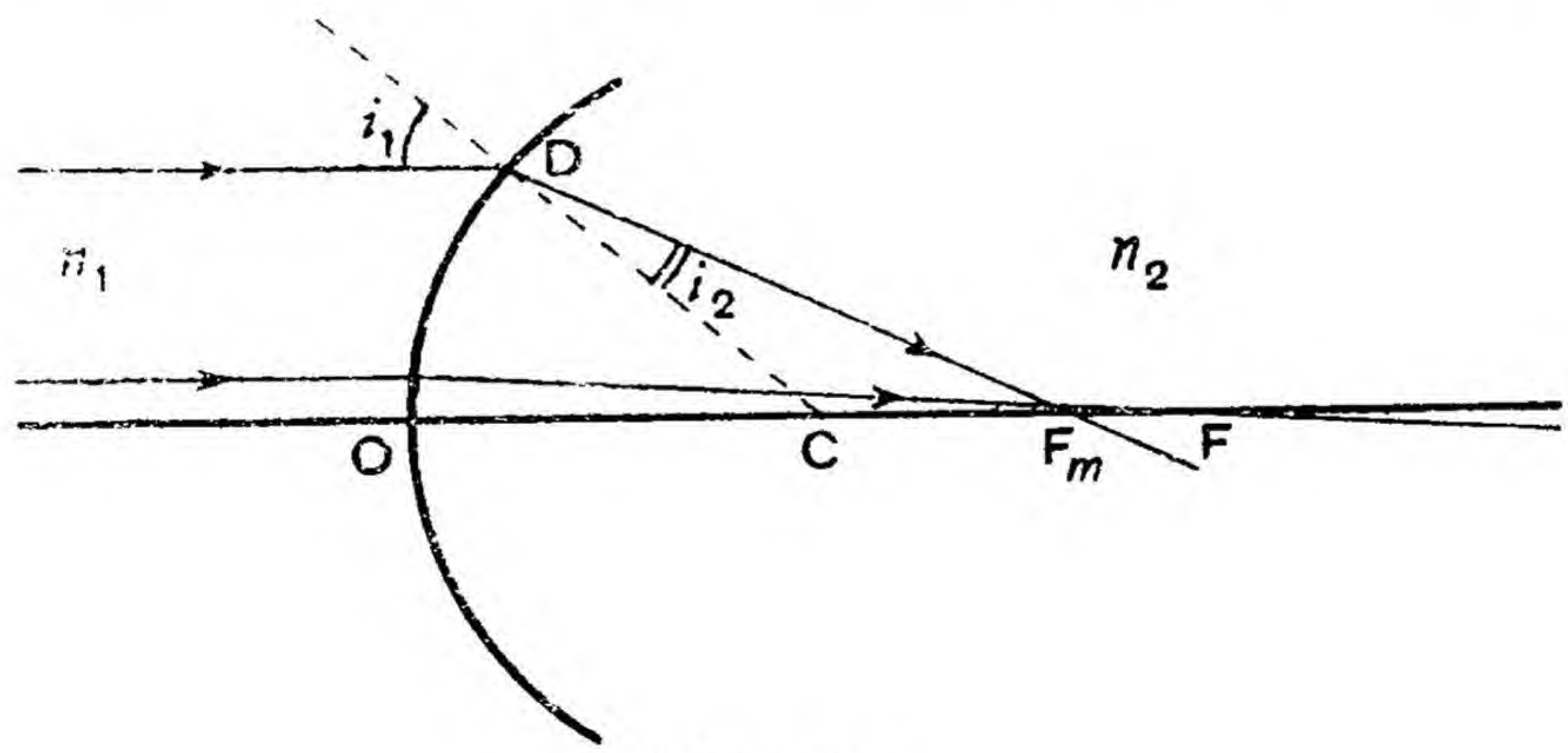


Fig. 65.

n_2 be greater than n_1 and let the surface be convex, as it is drawn. The paraxial rays come to a focus at the point F such that

$$\frac{n_2}{\text{OF}}=\frac{n_2-n_1}{\text{OC}}$$

as can be seen from equation (36) by putting $p=-\infty$, $\cos i_1=\cos i_2=1$ and $r=+$OC and solving for q, which equals $+$OF. But any marginal ray incident at an angle i_1 (Fig. 65) will cross the axis at the point F_m such that

$$\frac{n_2}{\text{DF}_m}=\frac{n_2 \cos i_2-n_1 \cos i_1}{\text{OC}}$$

Now, in the triangle CDF_m

$$\frac{CF_m}{\sin i_2}=\frac{DF_m}{\sin(\pi-i_1)}$$

$$\therefore\ CF_m=\frac{DF_m \sin i_2}{\sin i_1}$$

$$=DF_m\frac{n_1}{n_2}$$

by the law of refraction.

$$\therefore\ CF_m=\frac{n_1\,.\,OC}{n_2\cos i_2-n_1\cos i_1}$$

Also $$CF=OF-OC=\frac{n_2\,.\,OC}{n_2-n_1}-OC=\frac{n_1\,.\,OC}{n_2-n_1}$$

From the law of refraction

$$n_2\sin i_2=n_1\sin i_1$$

$$\therefore\ n_2^2\sin^2 i_2=n_1^2\sin^2 i_1$$

$$\therefore\ n_2^2(1-\cos^2 i_2)=n_1^2(1-\cos^2 i_1)$$

$$\therefore\ n_2^2\cos^2 i_2-n_1^2\cos^2 i_1=n_2^2-n_1^2$$

$$\therefore\ n_2\cos i_2-n_1\cos i_1=\frac{n_2^2-n_1^2}{n_2\cos i_2+n_1\cos i_1}$$

Since n_2 and n_1 are positive, $n_2\cos i_2+n_1\cos i_1$ decreases as i_1 and i_2 increase; therefore $n_2\cos i_2-n_1\cos i_1$ increases as i_1 and i_2 get greater. So it follows that $n_2\cos i_2-n_1\cos i_1$ is greater than n_2-n_1, its value when $i_1=i_2=0$. Therefore CF_m is less than CF. That is, in the case of refraction at one convex spherical surface from a less to a more dense medium, the focal length of marginal rays is numerically less than that for paraxial rays, or, **the marginal rays come to a focus at a point on the axis nearer to the pole of the surface than the paraxial rays.**

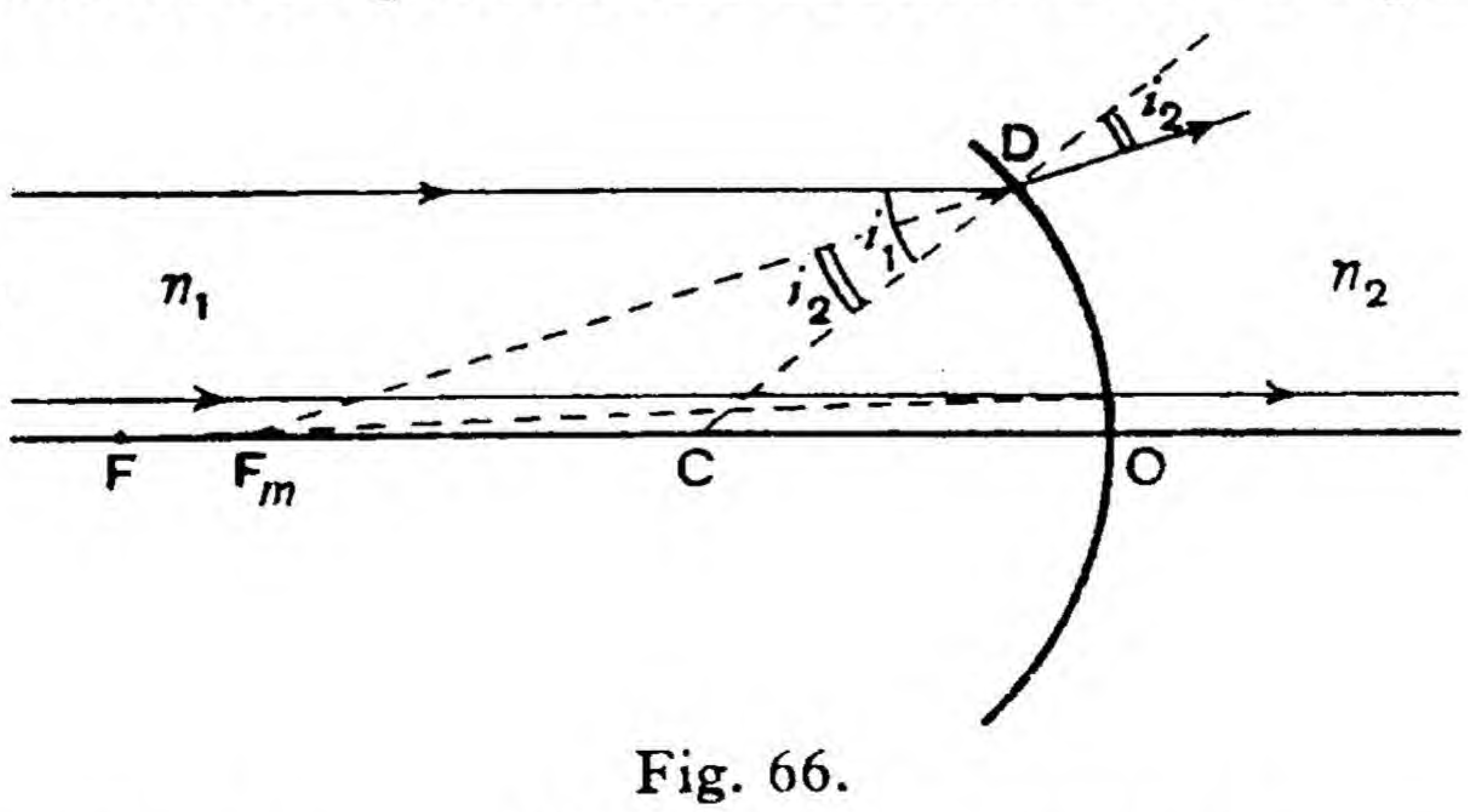


Fig. 66.

The reader will have no difficulty in deriving the same result from Figs. 66, 67, and 68 illustrating the remaining possible cases. In deriving these results he is advised to follow the method adopted in the first case and to convert the quantities q and r in equation (36), which are to be

allotted signs in accordance with the convention of Art. 10, into the corresponding distances DF_m and OC, which are pure numbers, and to work out the remainder of the proof in terms of these pure numbers. If any of these diagrams is rotated through 360° about the axis of the surface, the point D describes a circle on the surface which is known as a **zone** of the surface. So we have **Theorem 11 : all rays parallel to the axis of any single refracting spherical surface and striking the same zone come to a focus at the same point on the axis after refraction and the greater the radius of the zone the nearer is the point to the pole of the surface.** Let us now apply this theorem to a converging and diverging lens. We will consider a typical converging lens, such as a bi-convex lens, and let a beam of rays parallel to its axis fall on the lens. If the paraxial rays come to a focus at the point F on the axis of the lens (Fig. 69), the rays from the outermost zone of the lens will come to a focus at a point F_m nearer to the lens. For we have shown above that the marginal rays come to a focus nearer to the lens than the paraxial rays in

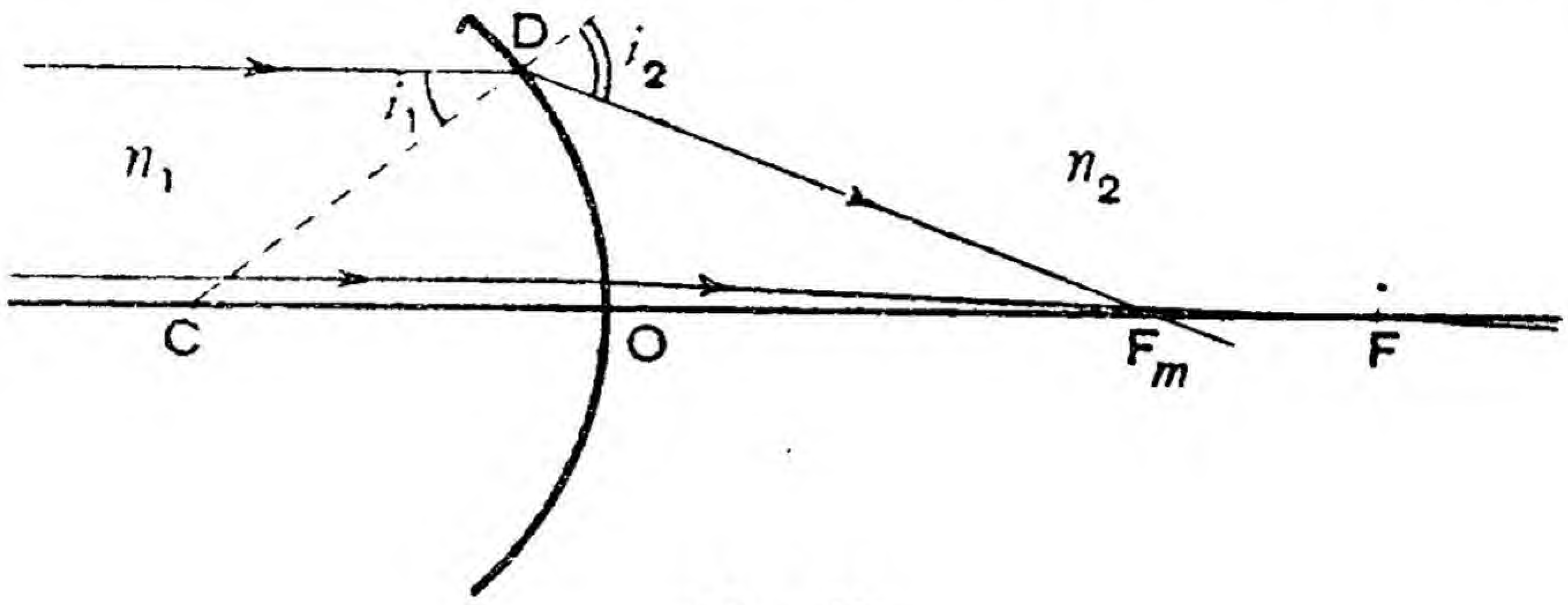


Fig. 67.

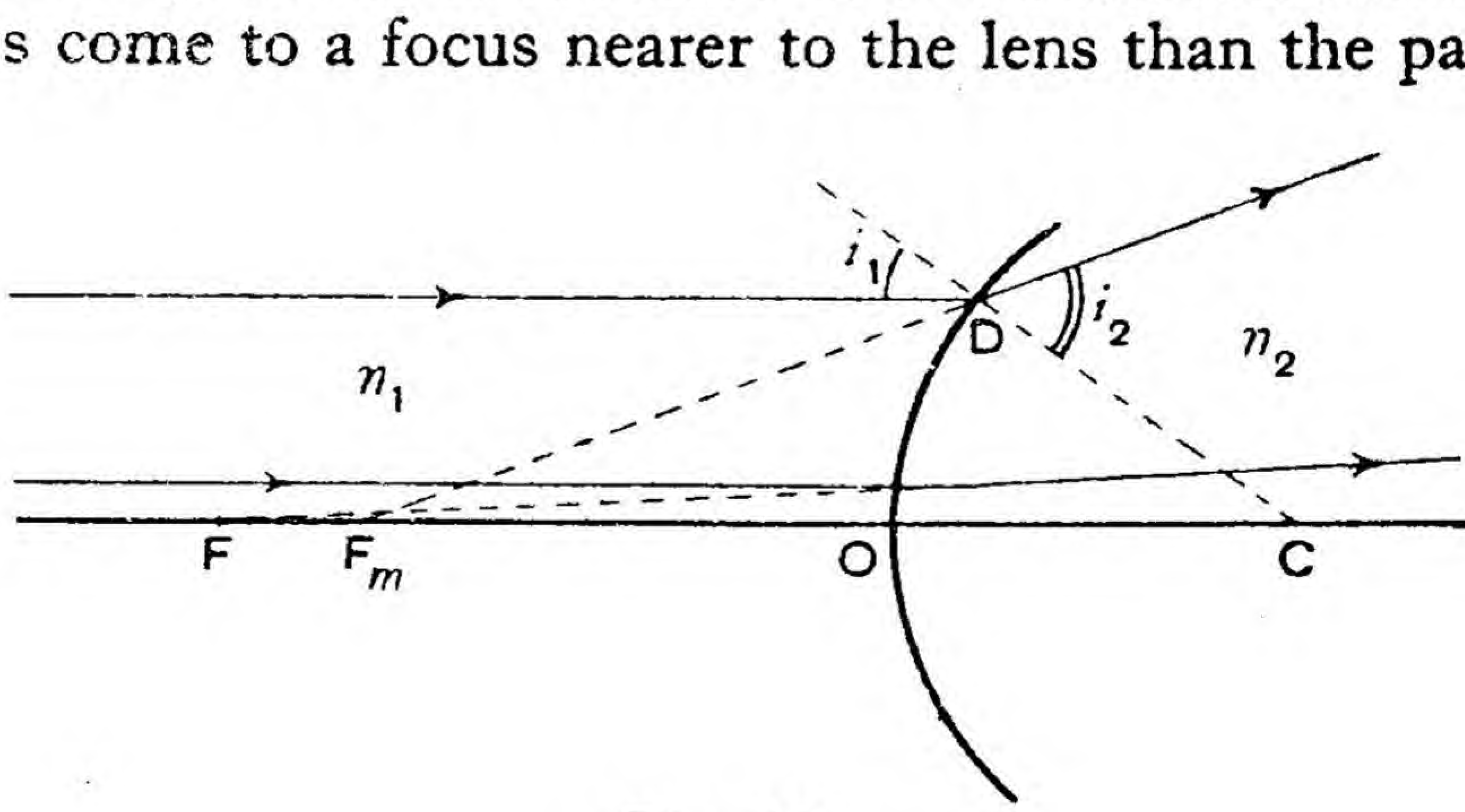


Fig. 68.

the refraction at the first surface and, since the second refraction is at a concave surface from a more to a less dense medium, this merely increases the difference in focal length due to the first refraction. The same result can be shown to be true for any type of converging lens. The fact that the rays from any given zone of the lens have a focal length, which varies with the radius of the zone, means that the image of a point object at infinity on the axis of the lens is in a different position according to the zone of the lens by which it is formed. Therefore, if the full aperture of

the lens is used to form an image, it is not possible to obtain a point image of a point object, since the rays from each zone of the lens come to a focus at points intermediate between the paraxial image and that formed by the marginal rays. It is evident that the best image is the circle of least confusion thrown on a screen placed in the position shown in Fig. 69, where the width of the beam is a minimum. Thus the sharpness or **definition** of the image of small objects close to the axis is poorer than if only paraxial rays are used. This defect of a lens, which is due to the spherical shape of its surfaces, is called **spherical aberration**. It is evident that it will

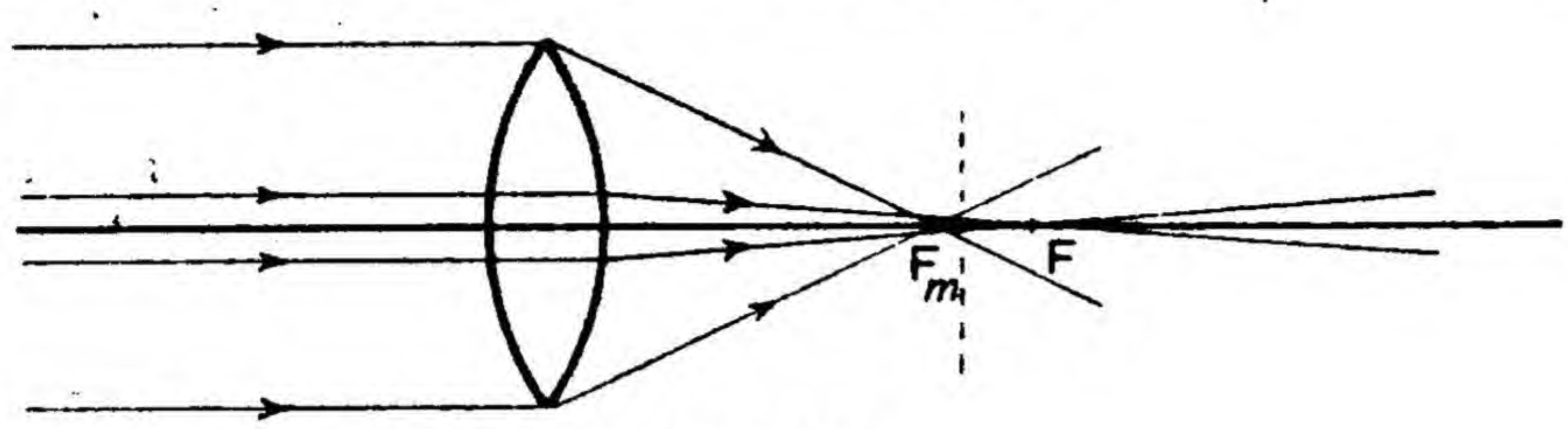


Fig. 69.

also be present in the case of a point object in *any* position on the axis of the lens. The quantity FF_m is called the **axial spherical aberration** of the given zone of the lens and is reckoned positive, when the focus of the marginal rays is on that side of the paraxial focus, from which the incident light is coming. The radius of the circle of least confusion is called the **lateral spherical aberration**. Both of these results, which are logical deductions from the axioms of geometrical optics, are verified by experiment; Plate I, Figs. 3 and 4, shows how the definition of the centre of the image becomes poorer as the aperture of the lens is increased.

Let us now consider a bi-concave lens as a typical diverging lens. Here again, if we consider a point object on the axis of the lens at infinity, the rays from the outermost zone come to a focus nearer to the lens than the paraxial rays. For we have already shown that this is the case for the refraction at the first concave surface and this effect is further increased by the second refraction from a more to a less dense medium at a convex surface. The same result can be shown to be true for any type of diverging lens. It is important to notice that the axial aberration is negative in this case, since the marginal focus is on the opposite side of the paraxial focus to that from which the incident light is coming. This suggests the possibility of eliminating or correcting spherical aberration by a suitable combination of converging and diverging lenses. So we have **Theorem 12 : rays parallel to the axis of any thin lens falling on an outer zone come to a focus at a point on the axis nearer to the lens than the paraxial rays.**

In order to investigate the conditions for producing a lens combination free from spherical aberration, it is necessary to find out how the axial spherical aberration depends on the radius of the zone of the lens, its focal length, the radii of curvature of its surfaces, and its refractive index.

There are two ways of doing this: the first is to derive a mathematical relation between these quantities working on the above lines. Unfortunately the expression, when it is obtained, is cumbersome and it is not easy to see from it how the axial spherical aberration varies with the above quantities. The second method is to take an actual numerical case and trace a ray through a given zone of the given lens by applying Snell's law to the refraction at each face and so to find the place where the ray crosses the axis. This is done for zones of various radii, lenses of various focal lengths, various radii of curvature of each face, the focal length being kept constant, and for various refractive indices. Then graphs are plotted showing how the axial spherical aberration depends on each of the above variables and from them some idea of how spherical aberration can be eliminated or reduced to a minimum can be obtained. We shall illustrate this method, which is a most important technique in the practical design of lenses, by finding both the axial and lateral spherical aberration for the telescopic objective, whose chromatic aberration was calculated in Art. 38.

45. CALCULATION OF SPHERICAL ABERRATION BY RAY TRACING

The way to trace a ray through an optical system can be understood by considering Fig. 70. The given incident ray ED is travelling in a medium of refractive index n_1 and strikes the spherical refracting surface DO,

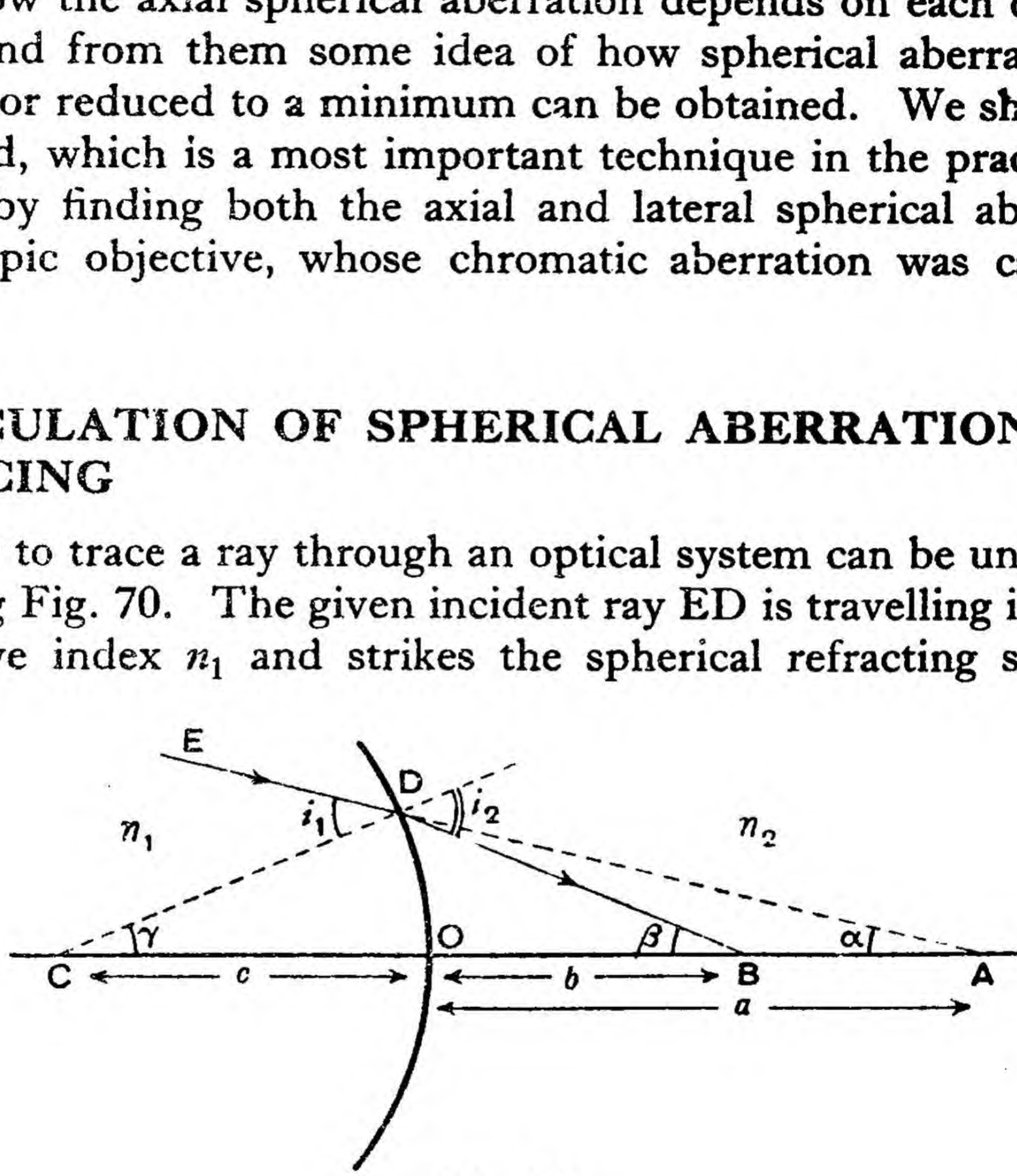


Fig. 70.

whose axis is CO, and is refracted along the path DB to cut the axis at B. The ray ED produced cuts the axis at A. The problem is given a, c, and α, to find b. This is done in the following steps: in the triangle ADC

$$\frac{\sin(\pi-i_1)}{a+c}=\frac{\sin\alpha}{c}$$

So i_1 is found from the equation

$$\sin i_1=\frac{a+c}{c}\sin\alpha$$

From the law of refraction we get i_2 from the equation

$$n_2 \sin i_2 = n_1 \sin i_1$$

Also we get γ from the equation

$$\gamma = i_1 - \alpha$$

and β from the equation

$$\beta = i_2 - \gamma$$

Finally in the triangle BCD

$$\frac{b+c}{\sin(\pi - i_2)} = \frac{c}{\sin\beta}$$

and so we get b from the equation

$$b = \frac{c \sin i_2}{\sin\beta} - c$$

This process can be carried on for each surface at which refraction occurs, provided that the centres of the various surfaces all lie on the same axis. Therefore the ray can be traced right through any refracting system, such as a combination of lenses, and we shall proceed to apply it to trace the

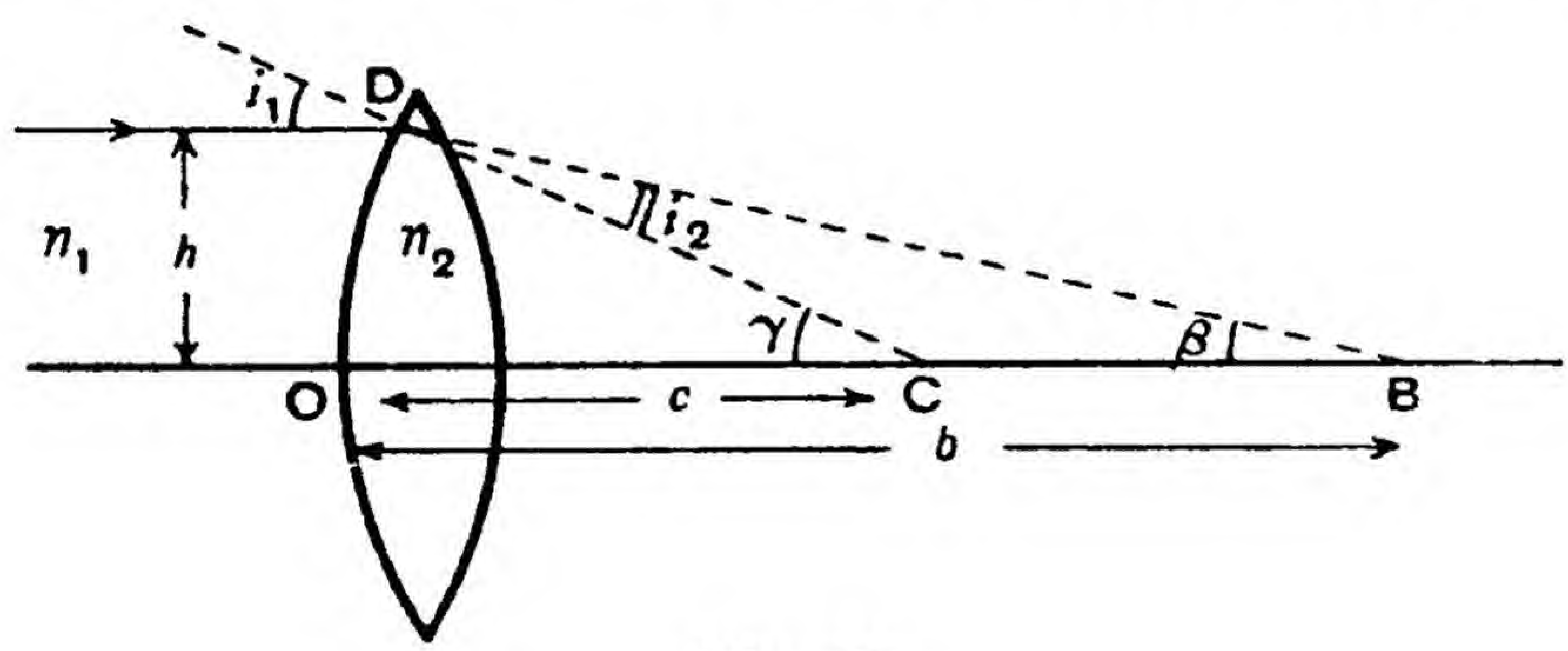


Fig. 71.

path of a ray through the above telescopic objective in air. The specification of the lens is as follows: a bi-convex lens, radius of curvature of each face 110 cm., axial thickness 2·00 cm., diameter 20·0 cm., refractive index of the glass for the D line 1·55. We shall trace the path of a ray parallel to the axis of the lens through the zone of radius 10 cm. The refraction at the first surface is illustrated in Fig. 71 and here we find i_1 from the equation

$$\sin i_1 = \sin\gamma = \frac{h}{c}$$

after which the determination of b proceeds as outlined above. We then proceed to the refraction at the second surface just as shown above and the complete calculation is shown in Table 6 on page 112.

So rays parallel to, and at a distance of 10 cm. from, the axis of this lens cross the axis after emerging from it at a distance of 98·1 cm. from the surface of the lens at which they emerge.

TABLE 6

Quantity.	First Surface.	Quantity.	Second Surface.
h	10·000 cm.	a	309·7−2·0=307·7 cm.
$\log h$	1·00000	$a+c$	417·7 cm.
$\log c$	2·04139	$\log (a+c)$	2·62087
$\log \sin i_1$	8·95861	$\log \sin a$	8·50920
		$\log c$	2·04139
$\log \frac{n_2}{n_1}$	0·19033	$\log \sin i_1$	9·08868
		$\log \sin i_2$	9·27901
$\log \sin i_2$	8·76828	i_1	7° 3′
i_1	5° 13′	i_2	10° 58′
i_2	3° 22′	γ	5° 12′
β	1° 51′	β	5° 46′
$\log \sin \beta$	8·50920	$\log \sin \beta$	9·00212
$\log (b-c)$	2·30047	$\log (b+c)$	2·31828
$b-c$	199·7 cm.	$b+c$	208·1 cm.
b	309·7 cm.	b	98·1 cm.

The point where paraxial rays come to a focus is found using the equation

$$\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{r}$$

for the refraction at each surface in turn. For the first surface we have $n_2=1{\cdot}55$, $n_1=1$, $u=-\infty$, $r=+110$ cm., from which

$$\frac{1{\cdot}55}{v}=\frac{1{\cdot}55-1{\cdot}00}{+110}$$

$$\therefore\ v=+310 \text{ cm.}$$

For the second surface we have $n_2=1{\cdot}00$, $n_1=1{\cdot}55$, $u=(310-2)=308{\cdot}0$ cm., and $r=-110$ cm., from which

$$\frac{1{\cdot}00}{v}-\frac{1{\cdot}55}{+308{\cdot}0}=\frac{1{\cdot}00-1{\cdot}55}{-110}$$

$$\therefore\ v=99{\cdot}7 \text{ cm.}$$

So the paraxial rays come to a focus 99·7 cm. from the surface of the lens at which the light emerges and therefore the longitudinal spherical aberration of the lens for the zone of 10 cm. radius is +1·6 cm.

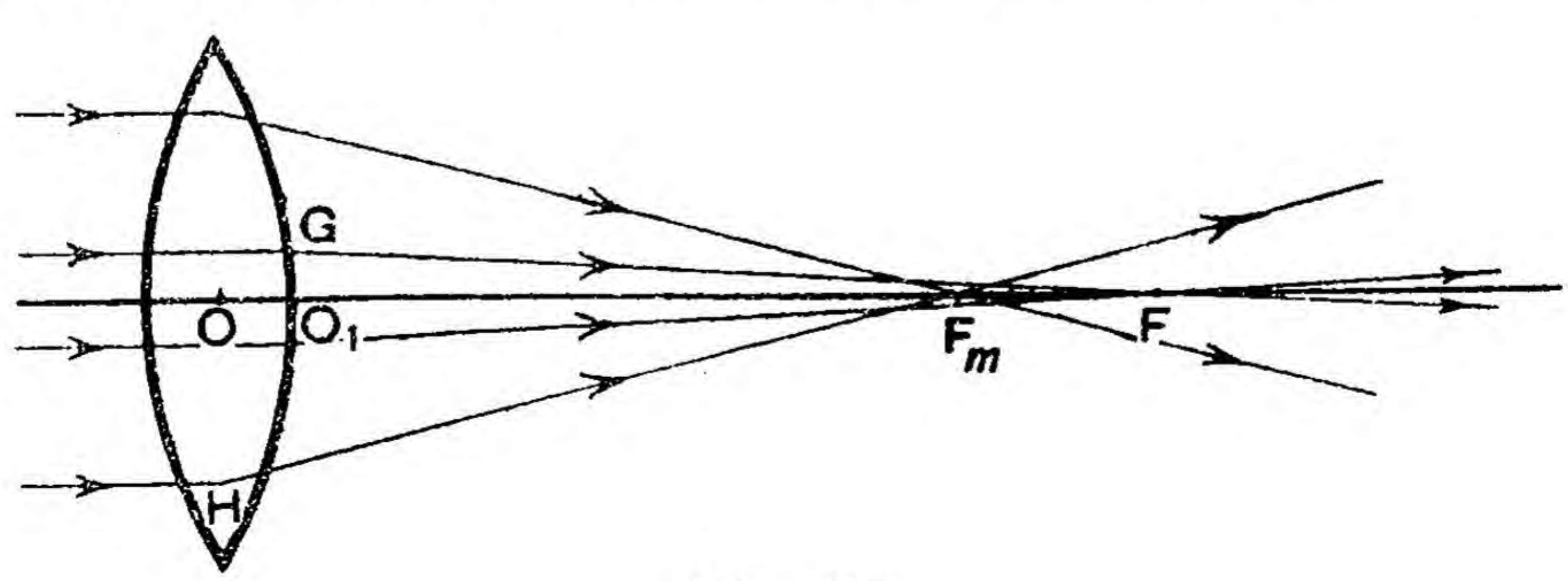


Fig. 72.

We shall now use this result to estimate the lateral spherical aberration of the lens. If the marginal rays come to a focus at F_m (Fig. 72), let

rays at a distance 2·5 cm. from the axis come to a focus at the paraxial focus F. We shall assume that the circle of least confusion has a radius PQ (Fig. 73) where P is the intersection of the rays HF_m and GF. This is a satisfactory way of fixing the size of this circle, for the ray GF is about the farthest one from the axis to satisfy paraxial conditions and rays

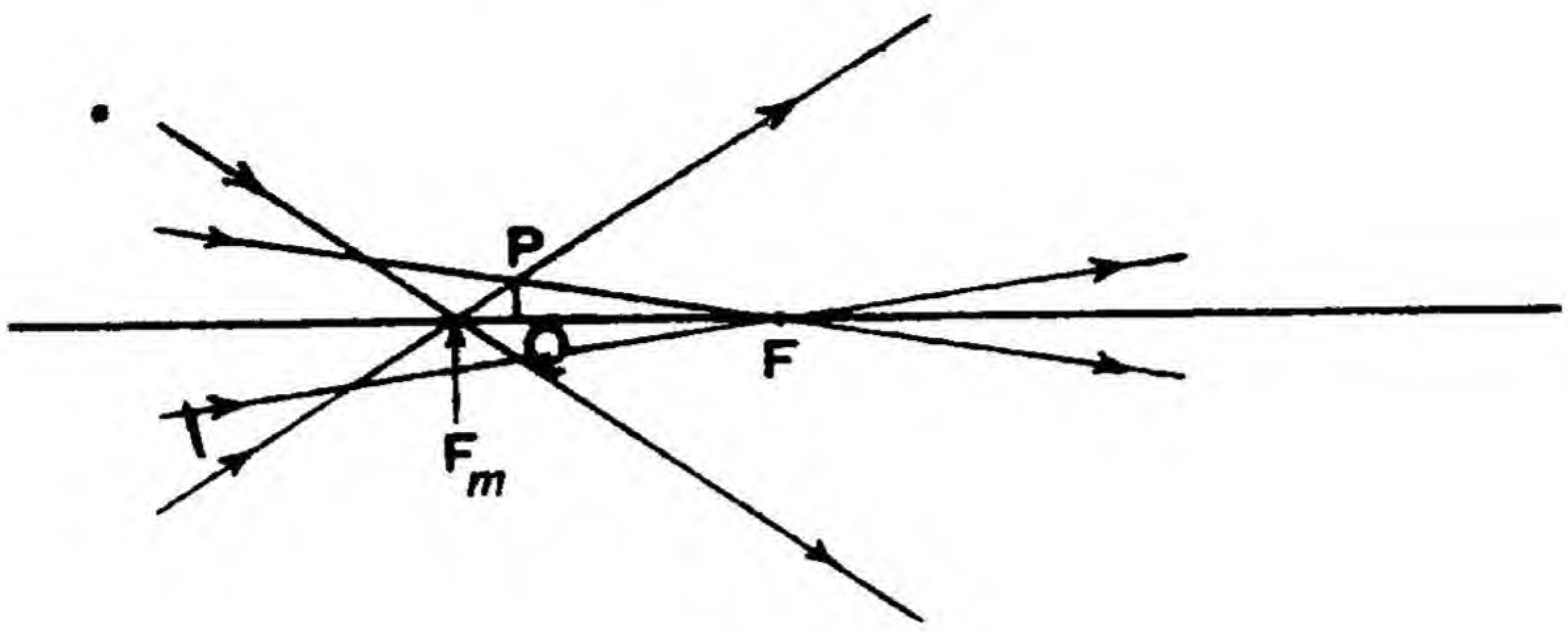


Fig. 73.

further out than this come to a focus nearer to the lens than F and so cut the cone of rays diverging from F_m in a circle of smaller radius than PQ. Now

$$\frac{PQ}{QF}=\frac{O_1G}{O_1F}=\frac{2{\cdot}5}{99{\cdot}7}$$

and $$\frac{PQ}{QF_m}=\frac{OH}{OF_m}=\frac{10}{98{\cdot}1+1{\cdot}0}=\frac{10}{99{\cdot}1}$$

But $$QF_m=1{\cdot}6-QF$$

$$\therefore \frac{PQ}{1{\cdot}6-QF}=\frac{10}{99{\cdot}1}$$

$$\therefore \frac{1{\cdot}6-QF}{QF}=\frac{2{\cdot}5}{99{\cdot}7}\times\frac{99{\cdot}1}{10}$$

whence QF=1·28 cm. and PQ=0·032 cm.

So we see that the best image, which can be formed by this lens, of a point object at infinity on its axis is a circle of 0·6 mm. in diameter and a similar result is true for a point object at any position on its axis. So we see to what extent this defect of spherical aberration spoils the definition of the image at points on the axis. It can, of course, be greatly decreased by diminishing the aperture of the lens by a stop placed on the same side of the lens as the object. But this remedy is not always available, as, for example, when a very faint star is to be photographed or an action photograph of some game has to be taken on a dull day, and other ways must then be sought.

The following results concerning axial spherical aberration have been established by the above method of trigonometrical ray tracing. The

8

graph of Fig. 74, which is taken from some results given in Hardy and Perrin's "Principles of Optics," gives the relation between the axial spherical aberration and the radius of the zone through which the ray passes and shows that the aberration is proportional to the square of this radius. The second important thing is, to see how the spherical aberration of a lens of given focal length varies as the radii of curvature of its faces are altered. This is called **bending** the lens. It is usual to express the aberration in millimetres for a lens of focal length 100 mm. for purposes of comparison, and the results shown in Table 7 are taken from Hardy and Perrin's "Principles of Optics," and refer to a lens of refractive index 1·51767 and axial thickness 20 mm., r_1 and r_2 being the radii of curvature of the faces, at which the light enters and leaves respectively. From these results a graph is plotted of spherical aberration against the curvature of the face at which the light enters (Fig. 75) and

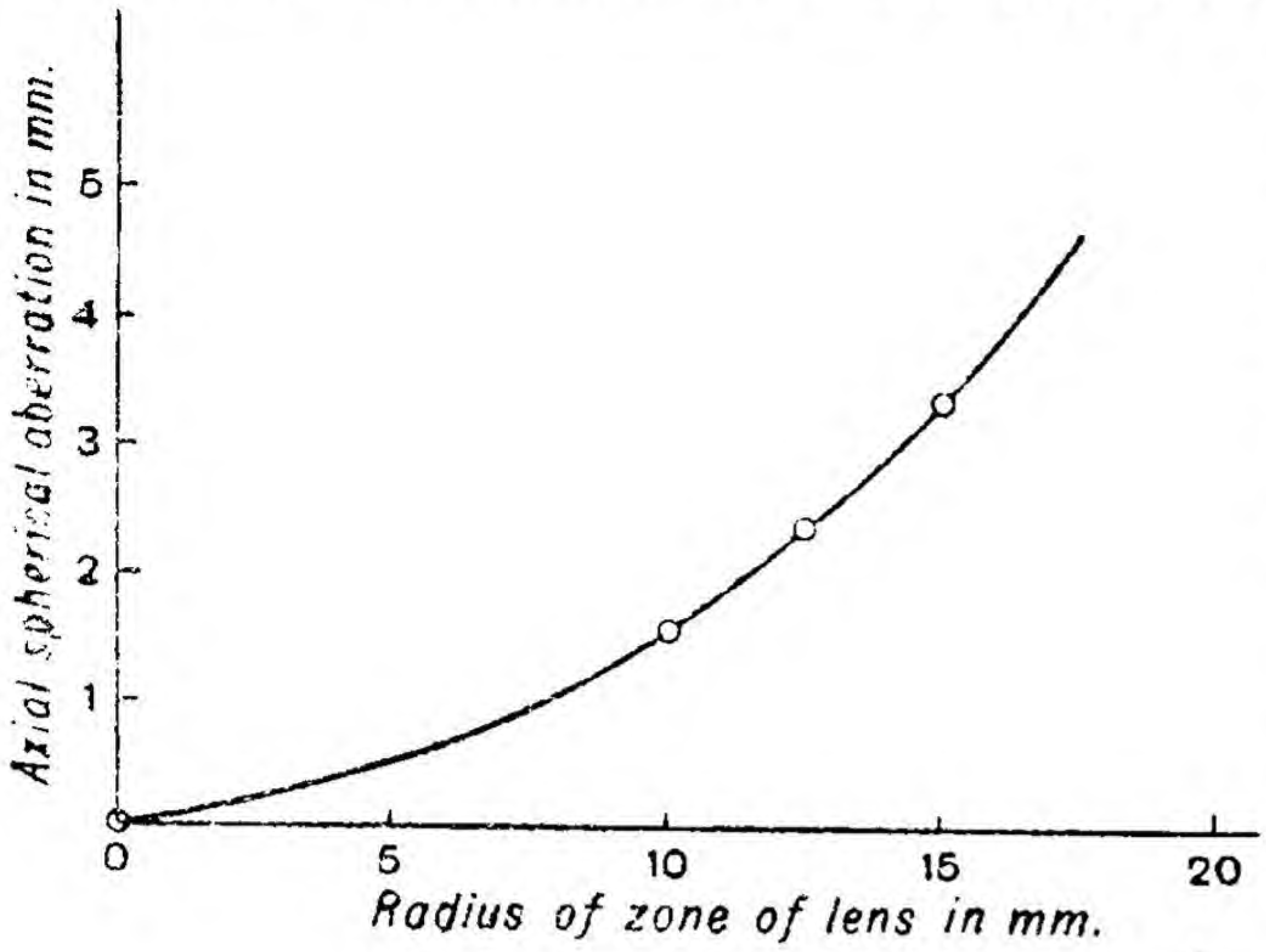


Fig. 74.

TABLE 7

r_1 mm.	r_2 mm.	$\frac{1}{r_1}$ in dioptres.	Spherical Aberration of a ray at 10 mm. in mm.
−150·0	−40 23	$\frac{1000}{-150} = -6{\cdot}67$	7·7
∞	−51·77	0·0	4·4
599·4	−56·02	1·67	3·7
425·2	−58·0	2·35	3·5
187·1	−68·97	5·35	2·5
100·0	−100·0	10·00	1·5
68·97	−187·1	14·50	1·0
58·00	−425·2	17·25	1·0
56·02	−599·4	17·85	1·0
51·77	∞	19·3	1·1
40·23	150·0	24·9	1·7

it will be seen that the aberration is large for lenses, in which the face at which the light enters is either concave or only slightly convex (Figs. 75 and 76). It reaches a minimum for lenses in which the face at which the light leaves is slightly convex or plane, after which it increases again as this faee hecomes more and more concave. The physical interpretation of this variation is that the spherical aberration is a minimum, when the total deviation of the ray is divided about equally between the two refractions (Fig. 77). The deviation of the angle of incidence from a paraxial

value is then equal in the two refractions and so the total spherical aberration produced is distributed equally between the two refractions. If the deviations are unequal, as in the case of the plano-convex lens, the greater spherical aberration is produced at the face where the deviation is the greater. The total spherical aberration is greater than when the deviations are equal, because the increase in spherical aberration at the

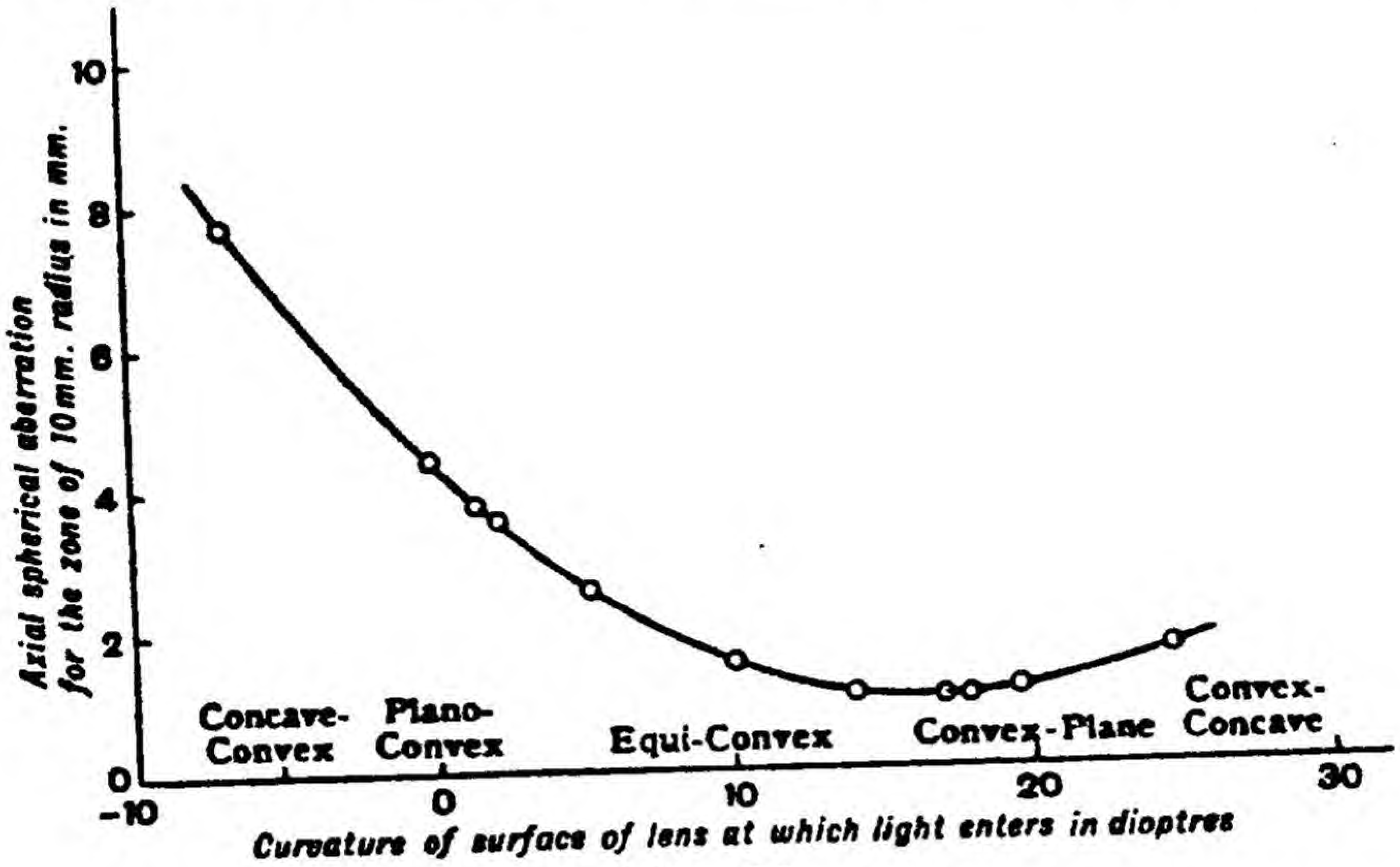


Fig. 75.

face where the deviation is increased is less than the decrease at the other face, since spherical aberration increases roughly as the square of the deviation. This way of regarding the condition for minimum spherical aberration also shows that the spherical aberration depends on the position of the object. For, if we consider a point object at a finite distance from a plano-convex lens, it is evident that this distance can be so arranged

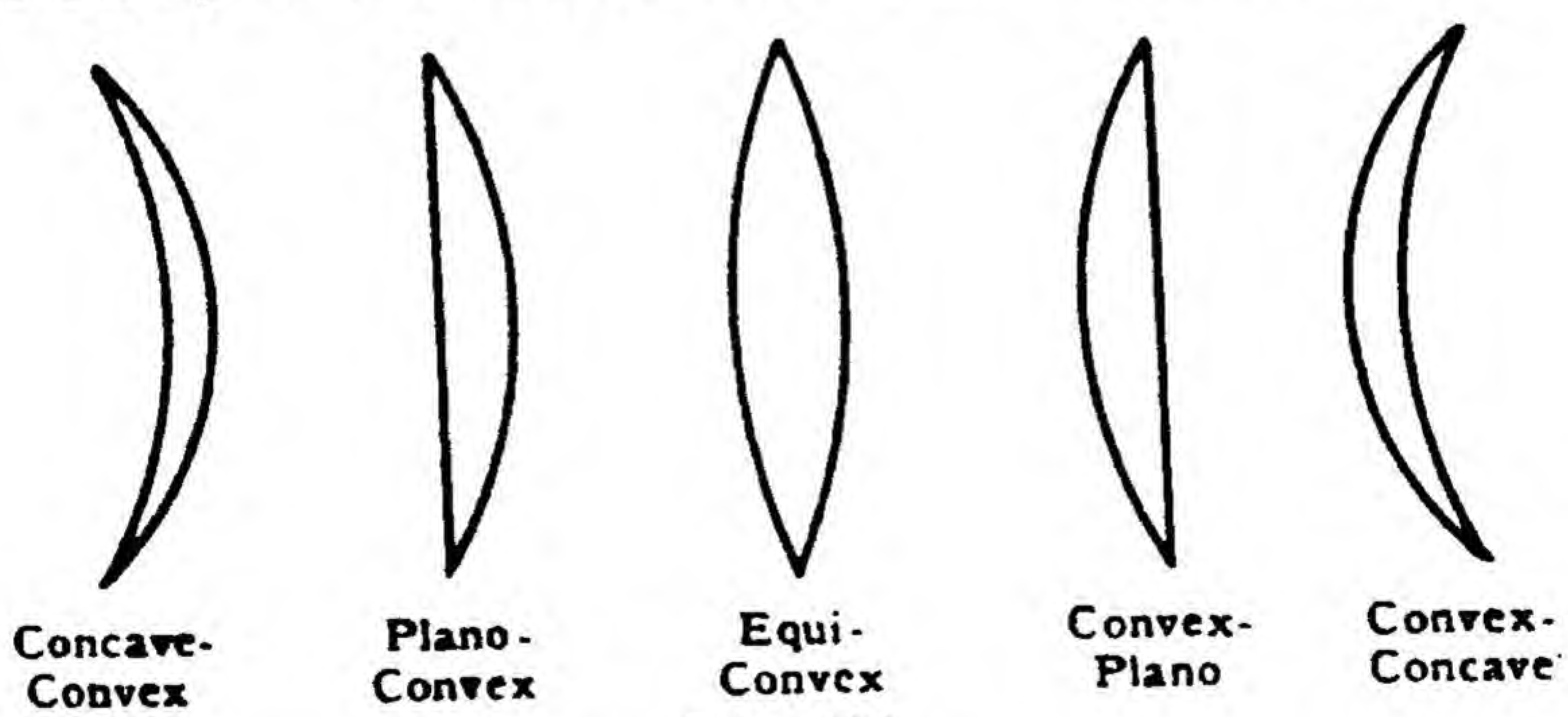


Fig. 76.

that the deviations at each face are equal and then the aberration is a minimum for an object in this position, whereas we have just seen that it is very large for an object at infinity. This result is verified by ray tracing, from which it can also be shown that the aberration is inversely proportional to the cube of the focal length of the lens. A relation between aberration and the refractive index of the lens can also be worked out, but it does not exhibit any simple features.

These results suggest the principle underlying the design of lens combinations free from spherical aberration. Let us suppose that we wish

to make a combination which is converging, then a converging lens must be combined with a diverging lens of longer focal length. If the two lenses were of the same shape, the aberration of the diverging lens would be less than that of the converging one and so, although it is in the opposite direction, it would not cancel it out. But if the shape of the converging lens is chosen so as to give it minimum aberration, then it may be possible to choose that of the other component so as to produce an equal and opposite aberration, and so the combination will be free from spherical aberration. The precise shapes of the two lenses have to be found by ray tracing, and it must be emphasised that the compensation is only perfect for one position of the object. Consequently the lens cannot be designed, until the purpose for which it is to be used has been specified. This account of ray tracing has been included to give the reader a glimpse of the methods used by technical opticians and lens designers, so that there is no necessity for him to study it deeply unless he contemplates taking up practical optics.

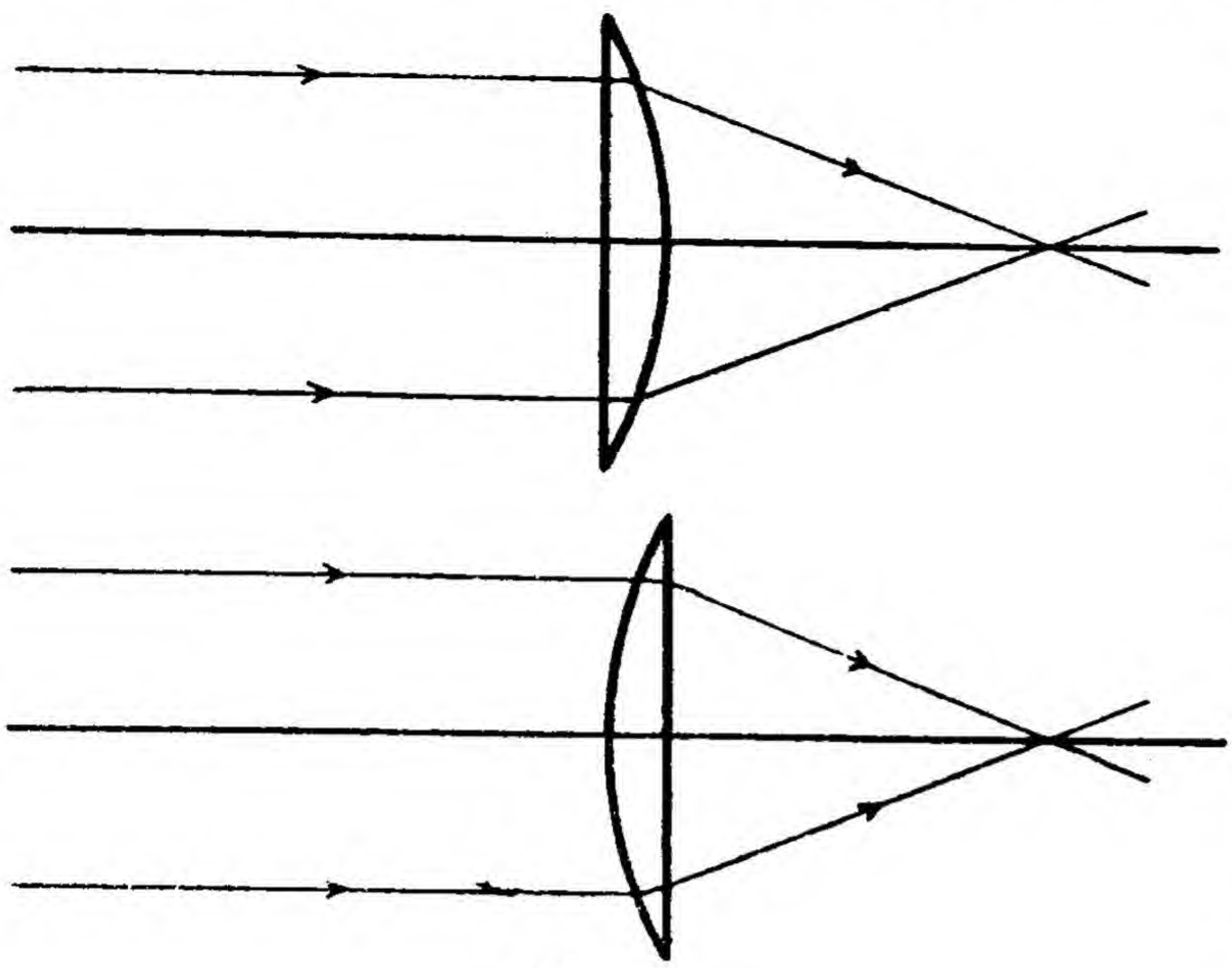


Fig. 77.

46. APLANATIC SURFACES

This account of spherical aberration will be concluded with the one case in which a spherical surface forms a geometrical point image of a

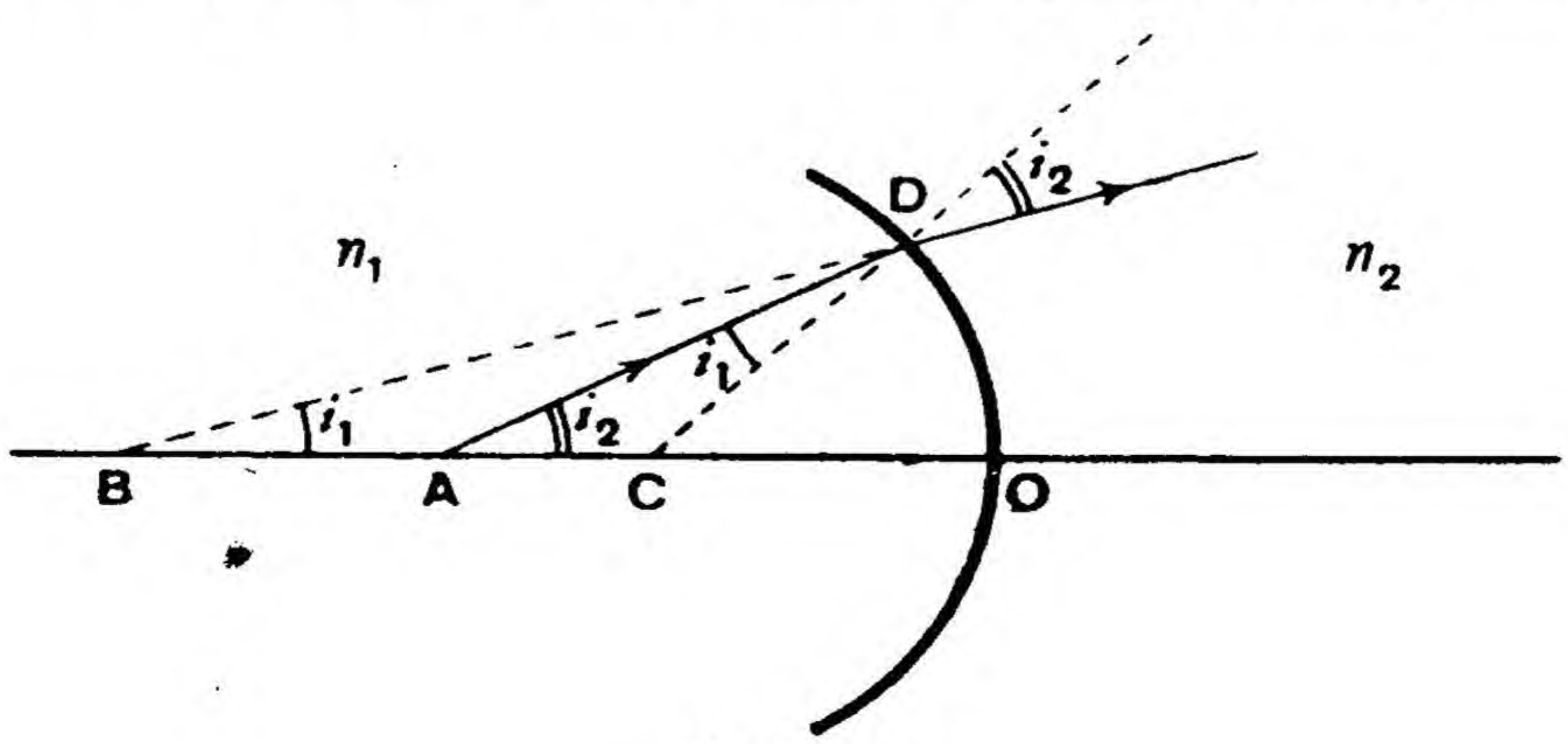


Fig. 78.

point object on its axis, however big the aperture of the surface. Let *any* ray AD from a point object A (Fig. 78) on the axis of a spherical refracting surface centre C, pole O, be refracted from the medium of

refractive index n_1 into the medium of *smaller* refractive index n_2, so that it appears to have come from the point B on the axis. If $AC=\frac{n_2}{n_1}.OC$, in the triangle ADC we have

$$\sin D\hat{A}C=\frac{DC.\sin i_1}{\frac{n_2}{n_1}.OC}$$

$$\therefore \sin D\hat{A}C=\frac{n_1}{n_2}.\sin i_1=\sin i_2, \text{ by the law of refraction.}$$

$$\therefore \angle D\hat{A}C=i_2$$

$$\therefore \angle D\hat{B}A=\angle D\hat{A}C-\angle A\hat{D}B$$
$$=i_2-(i_2-i_1)$$
$$=i_1$$

Therefore the triangles ADC and DBC are similar.

$$\therefore \frac{BC}{CD}=\frac{DC}{CA}$$

$$\therefore BC=\frac{CD^2}{CA}=\frac{OC^2}{\frac{n_2}{n_1}.OC}$$

$$\therefore BC=\frac{n_1}{n_2}.OC$$

and is independent of the angle i_2, which AD makes with the axis of the surface. Therefore *all* the rays from A, *however large an angle they make*

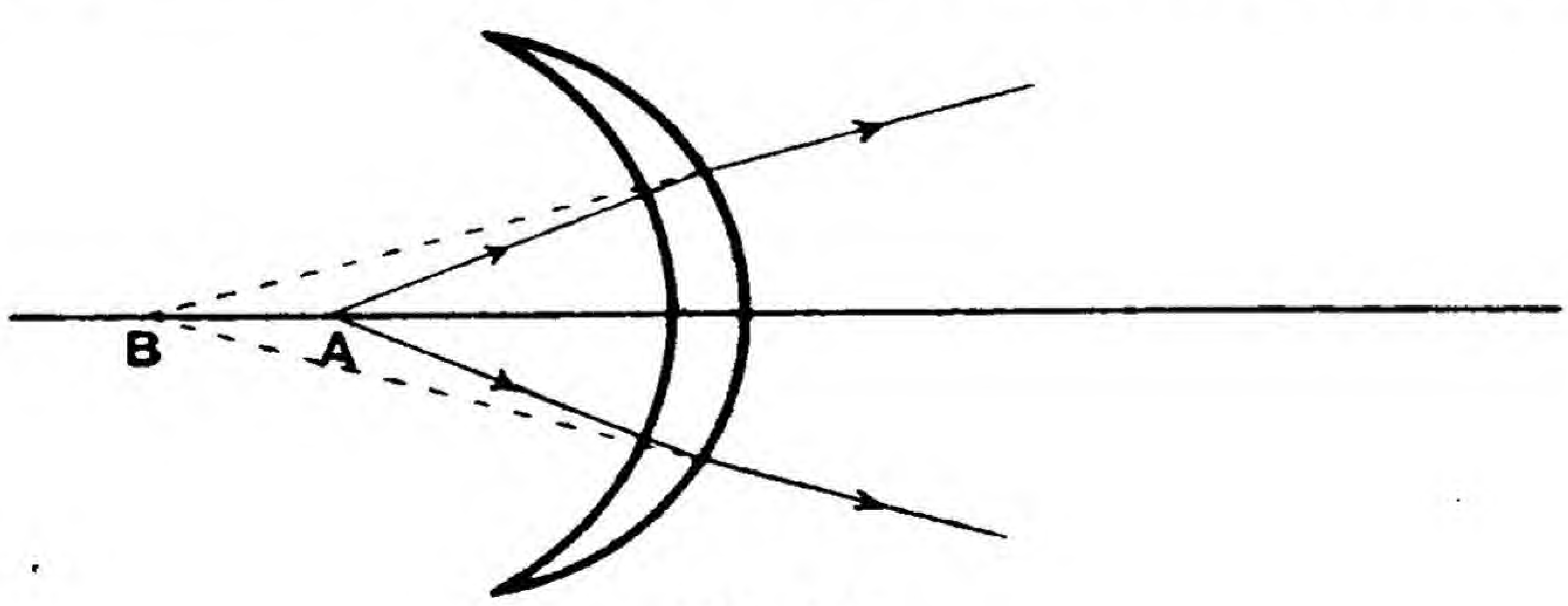


Fig. 79.

with the axis, appear to diverge from B after emerging from the spherical boundary between the two media, and so there is no spherical aberration for this one position of the object. Any surface which satisfies both this condition and the sine condition referred to in Art. 49 is called an aplanatic surface. Such surfaces are made use of in microscopic objectives, where it is essential that as wide a pencil as possible from each point of the object should enter the objective, otherwise the greatly magnified image will be too faint to be seen distinctly. The objective is designed so that the

centre of curvature A of the face at which the light enters coincides with the aplanatic point for the other face (Fig. 79). If a point object is placed at this point, the rays from it enter the lens normally and therefore suffer no refraction at the first face, and the refraction at the second surface is without any spherical aberration. Therefore the wide angle pencil from the object at A is changed by the lens into a narrower pencil diverging from B, and the remaining components of the objective can be corrected for spherical aberration along the lines indicated above, which would have been insufficient for the initial wide angle pencil.

47. POINTS OFF THE AXIS

We have so far considered only points on the axis of the lens, and it is now time to see what happens, if a lens receives rays from a point object so far off its axis, that paraxial conditions are violated even for a lens of small aperture. The general nature of the problem can be understood from Fig. 80, in which B is the point image of a point object A on the

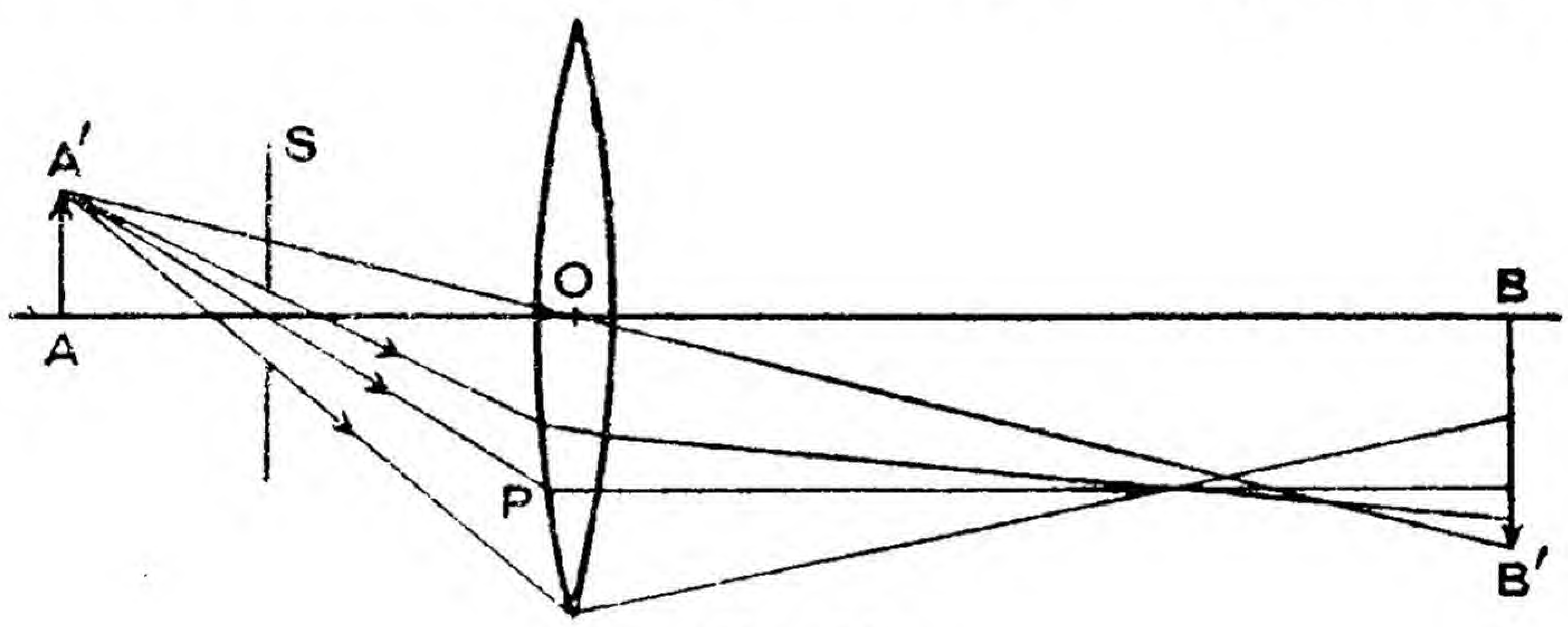


Fig. 80.

axis of a converging lens, being formed entirely by paraxial rays. If the diagram is rotated through a small angle about the point O as centre, then A describes the short line AA′ perpendicular to the axis of the lens and B another short line BB′ and we have already seen (Art. 13) that B′ is the place where the image would be, if the rays forming it satisfied paraxial conditions. In future we shall call this point the paraxial image of the object A′. We must now see where the actual image of A will be formed and whether a point image can be obtained, if paraxial conditions are not satisfied. To consider a particular but typical case, let us suppose that a stop S is placed between the lens and the object, which therefore sends a pencil of rays on to the lower part of the lens. At what point or points will these rays come to a focus and how will these points be related to the paraxial image B′? The ray A′P passing through the middle of the stop is called the principal ray of the pencil. We can get a general idea of the effects to be expected by applying the results we have just obtained for axial spherical aberration to this case, A′OB′ being the axis to which the axial aberration is referred. We shall call it the **auxiliary axis** in future. Since the axial spherical aberration of a ray is proportional to

the square of the radius of the zone of the lens through which it passes, this aberration will be much larger for the rays from A′ than those from A, since they pass through zones of much larger radius. *Hence defects will appear in the images of points off the axis, even when the aperture of the part of the lens utilised is so small that the spherical aberration of points on the axis is negligible* and the remaining conclusions apply only to small apertures. We see that the three rays drawn cross the auxiliary axis at different points, since they have come from different zones of the lens, and it is evident that the emergent pencil will be asymmetrical relative to the principal ray on account of the asymmetrical position of the object. Its precise nature is complex and can be investigated either mathematically or by ray tracing. We shall not go into any details, but merely try to explain clearly the results of the investigations.

The mathematical treatment leads to an expression containing four different terms, each of which corresponds to a particular defect in the image formed by the lens, and these defects will now be considered in turn. The first three, **spherical aberration, coma, and astigmatism** affect the definition of the image, while the last one, **distortion**, affects its position. It is important to realise that any defect is not significant until the previous ones have been corrected out. The reason for this is that, if the previous defects have not been removed, the image suffers from a combination of all the remaining defects, which often cannot be explained in simple terms.

48. SPHERICAL ABERRATION

The first term in the expression indicates that the marginal rays of the pencil passing through the stop come to a focus on the principal ray nearer to the lens than the rays very close to the principal ray and that the distance between these two foci is proportional to the square of the distance of the marginal ray from the principal ray, as they pass through the stop. In fact, this defect has just the properties of spherical aberration, and so we see that, if this defect is present for points on the axis, it will be present to the same extent for points off the axis too. If it has been eliminated or reduced to a minimum for axial points, it will also be absent for points off the axis.

49. COMA

The second term in the expression gives rise to a defect known as coma, which is serious for points close to the axis of a lens and is due to the asymmetrical position of the object relative to the lens. We can best appreciate the cause of this by considering the formation of the image of a small object AA′ by a pencil, whose principal ray passes through the centre of the lens, assuming that spherical aberration has been eliminated. (Fig. 81). Precisely similar effects are produced when the principal ray is excentric. The paraxial images of A and A′ are at B and

B′ respectively, but the rays 11 from A′ passing through an outer zone of the lens come to a focus at B_1', while those farther out still come to a focus at B_2'. Therefore the paraxial rays form the image BB′, the rays from an inner zone form the image BB_1', while those from the outer zone form the image BB_2'; the fact is that the lateral magnification of the lens is different for different zones, decreasing as we go out-

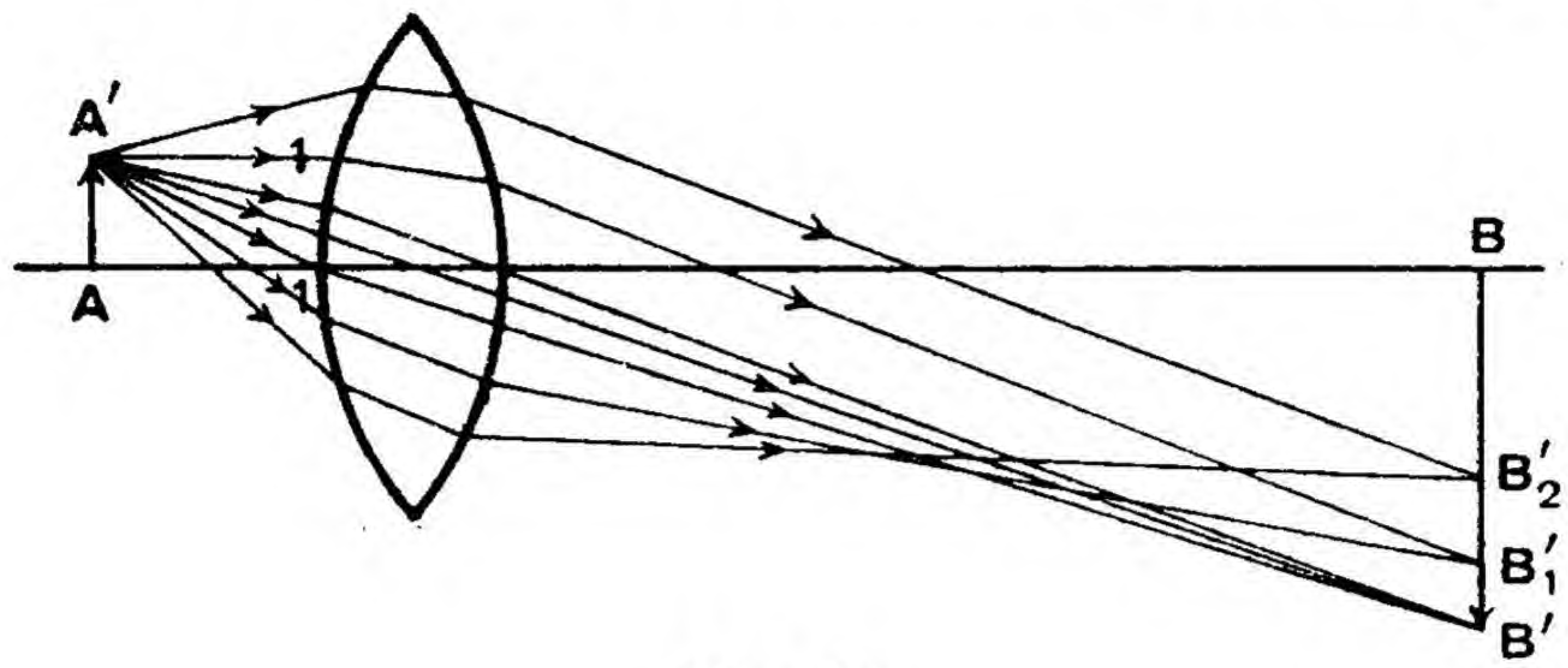


Fig. 81.

wards. But so far we have only considered the rays in the plane containing the point A′ and the axis; when we come to consider the **skew rays**, which do not lie in this plane, the nature of the image is even more complicated and is deduced from Fig. 82, which gives a view of the lens looking at one of its faces. The centre of the lens is at O and the zone of the lens is marked off through which the two rays marked 11 in the previous

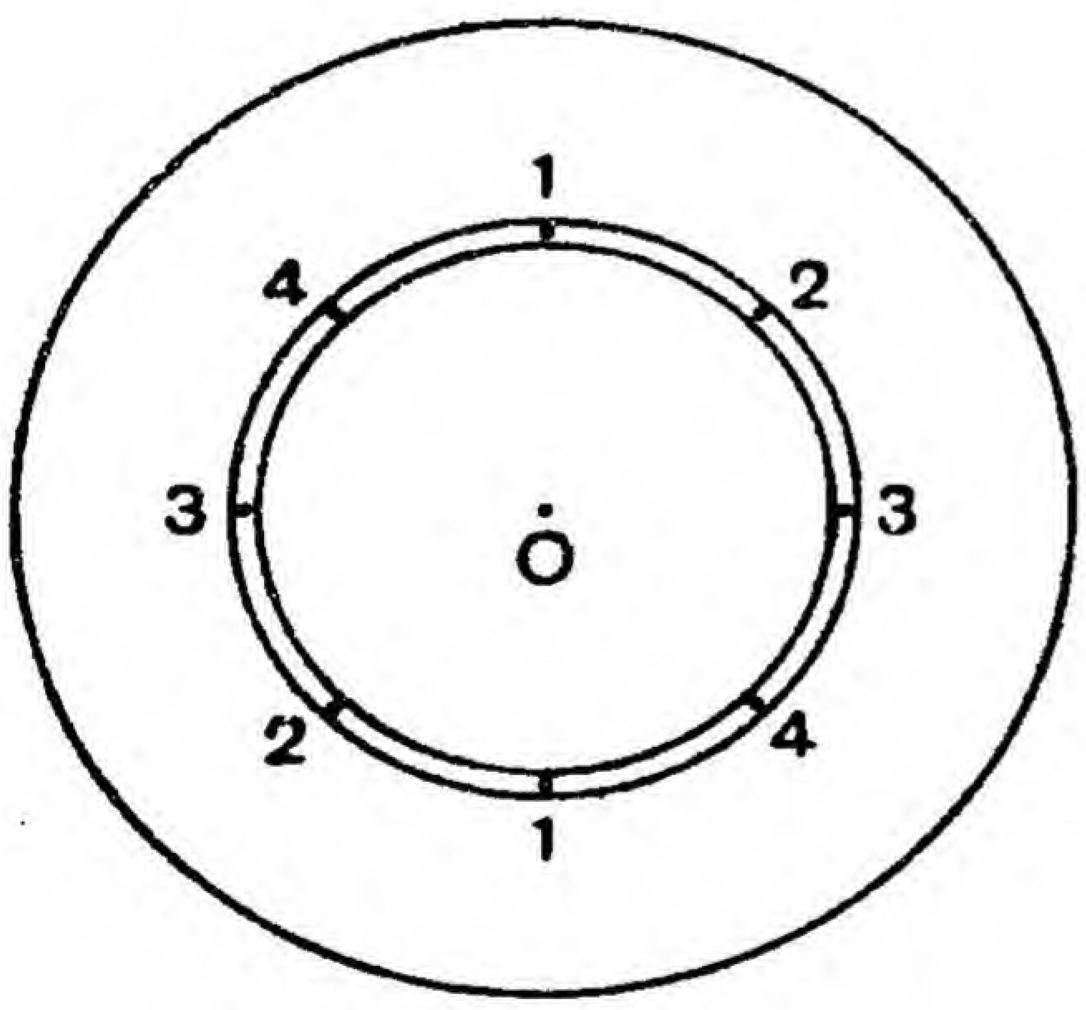


Fig. 82..

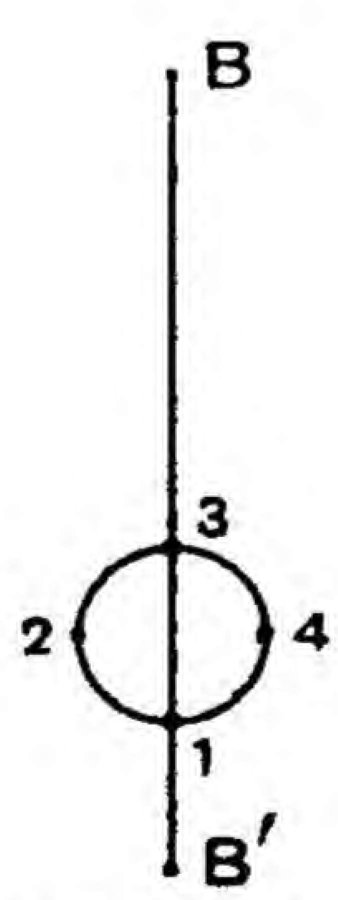


Fig. 83.

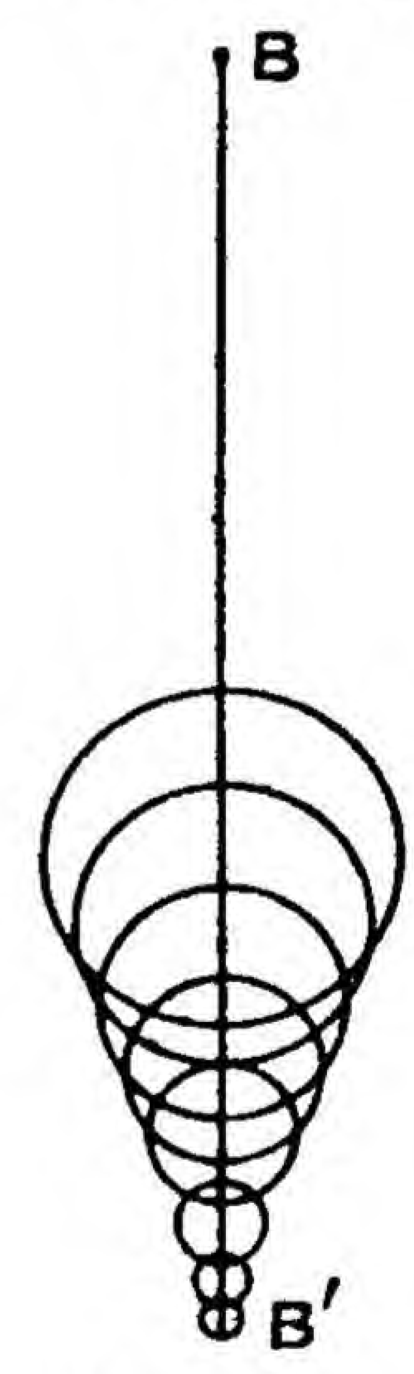


Fig. 84.

figure pass. They come to a focus at the point marked 1 in Fig. 83, the rays 22 coming to a focus at the point 2 and so on, the point 1 in Fig. 83 coinciding with the point B_1' in Fig. 81. It follows, therefore, that the rays from A′ passing through this zone of the lens come to a focus in the circle 1234 and the radius of the circle increases as that of the zone increases. So the complete image of the point A′ formed by the whole lens is a set of circles of increasing radius, starting at the point B′ and going upwards towards B as shown in Fig. 84. The intensity of illumination is greatest at B′ and decreases as we go upwards towards B_2', and so

PLATE II

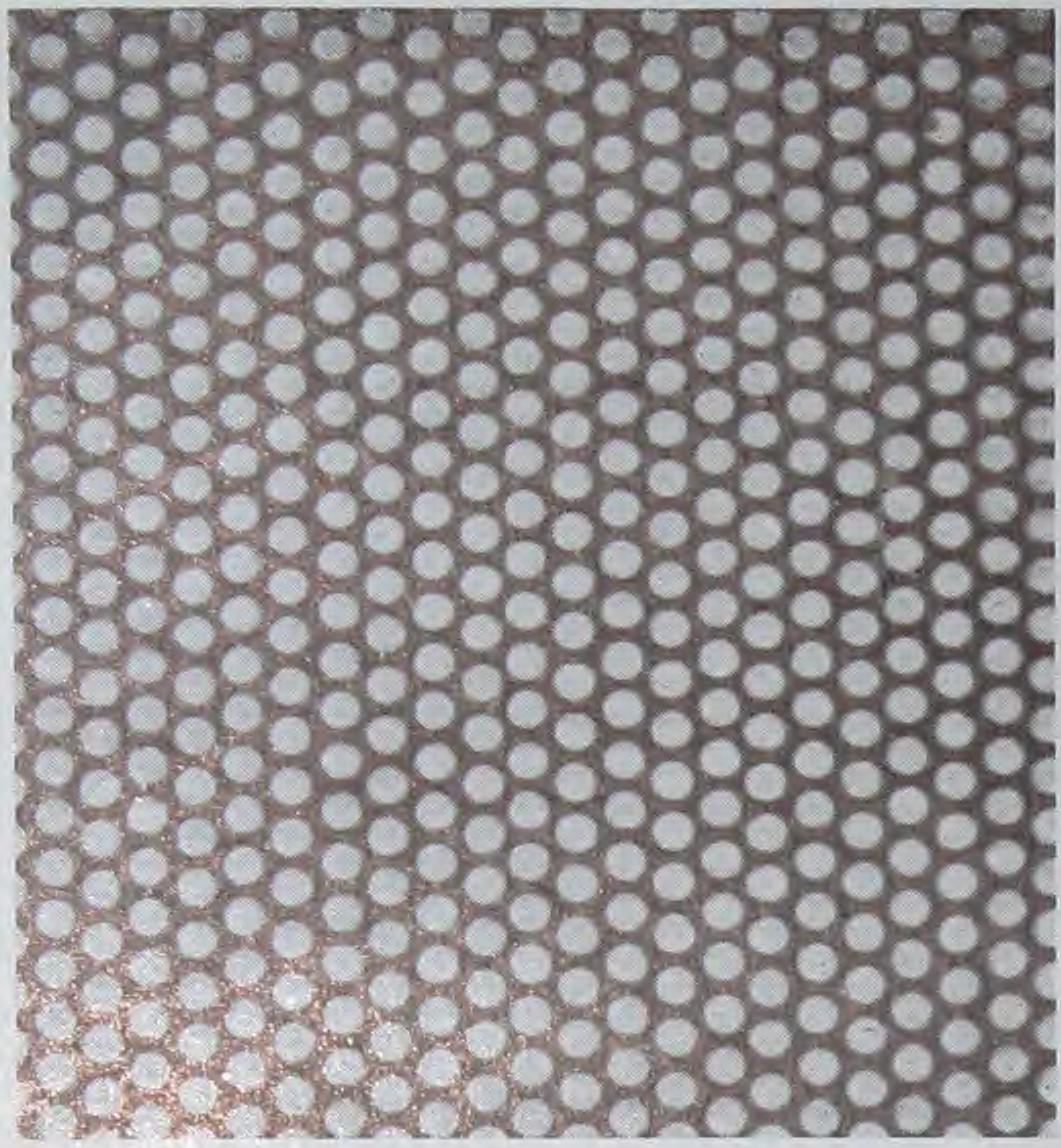


Fig. 1. Image of a gauze formed by a plano-convex lens with all the deviation occurring at the convex face, producing large spherical alteration and poor definition.

(*J. W. Mitchell*)

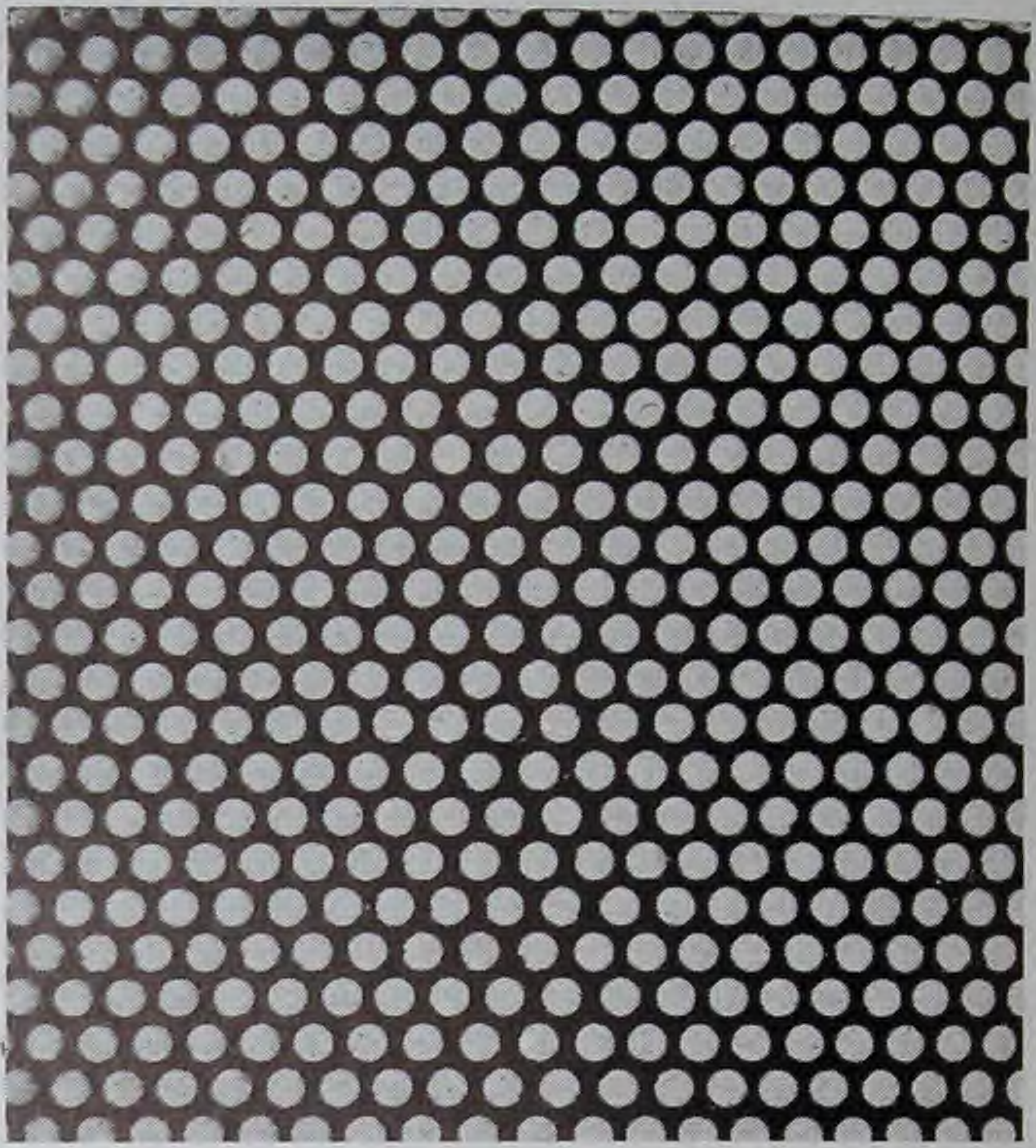


Fig. 2. Image of the same gauze formed by the same lens with the deviation divided between the two faces, producing minimum spherical alteration and good definition.

(*J. W. Mitchell*)

Fig. 3. The image of a pin-hole on the axis of a convex lens.

(*J. W. Cottingham*)

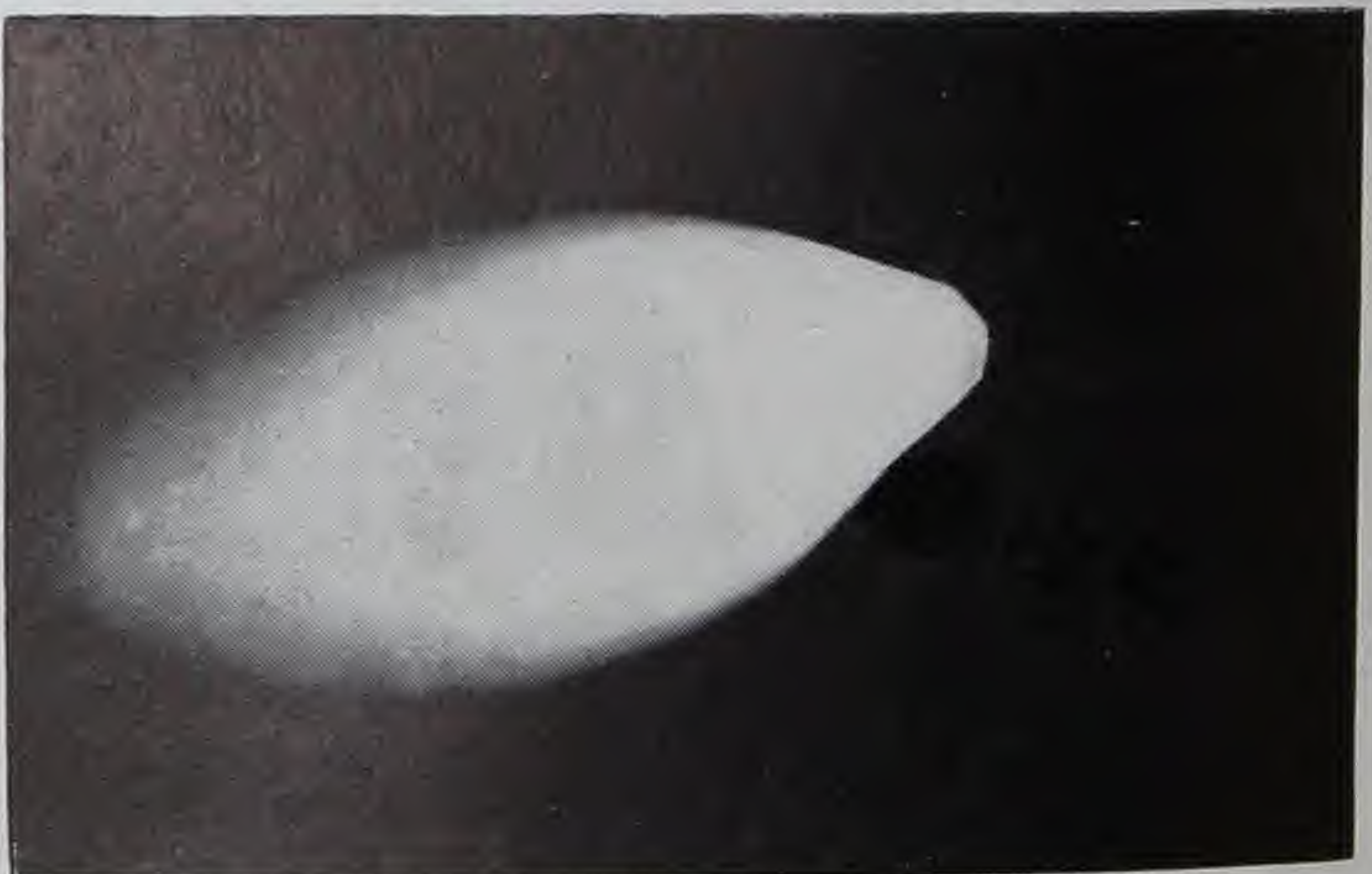


Fig. 4. The image of the same pin-hole to one side of the ax.s of the same lens showing coma.

(*J. W. Cottingham*)

the appearance of the image is similar to that of a comet with its tail, from which the name coma is derived. It is evident that this defect will spoil the definition, and it is particularly undesirable on account of the asymmetrical form it gives to the image, which makes it quite impossible to set the intersection of the cross-wires of a microscope, for example, on the centre of the image. It can be demonstrated by forming a magnified image of a small hole placed on the axis of an achromatic plano-convex lens used with the plane side towards the object, so that spherical aberration is a minimum. When the centre of the hole is on the axis of the lens, the definition is at its best, but if the lens is turned about a vertical axis through two or three degrees it will be seen that the definition has deteriorated, and if the lens is turned through some ten degrees, the appearance of the image is as shown in Plate II, Fig. 4, the bright part of the image being due to the central rays and the remainder to those which have passed through the outer zones of the lens, as can easily be verified by stopping out these two sets of rays in turn.

As this defect occurs in points close to the axis of a lens of large aperture, it is particularly serious in the case of microscopic objectives.

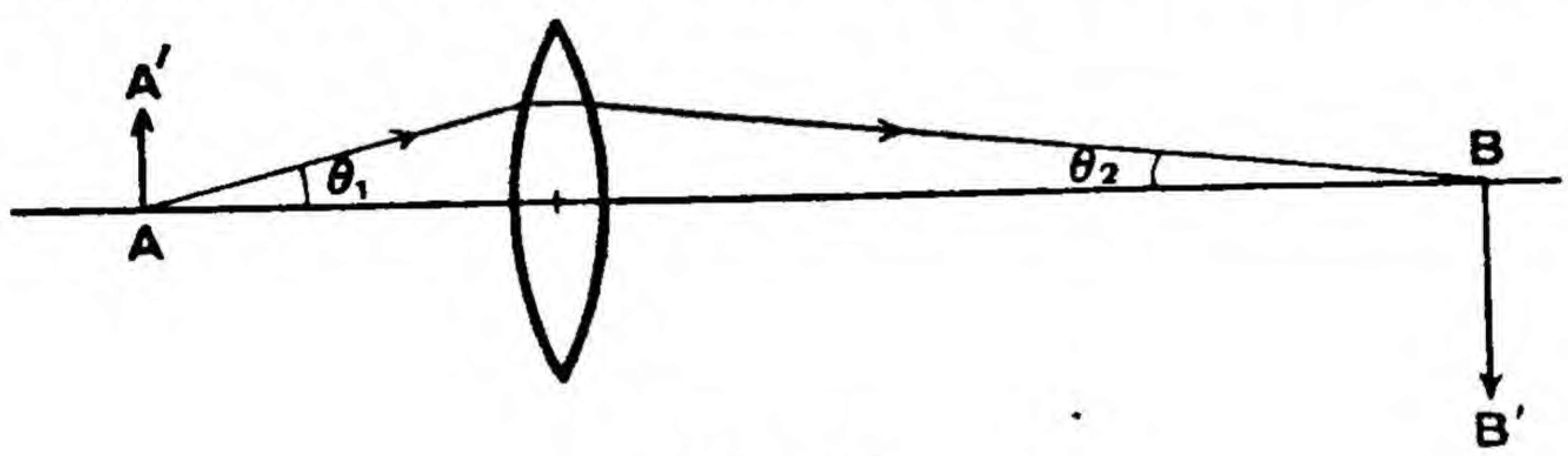


Fig. 85.

It has to be corrected by designing the lens so that the magnification is the same for all zones of the lens. If BB′ is the image of an object AA′ formed by a lens (Fig. 85) and θ_1 and θ_2 are the inclinations of a ray from A to the axis before and after emerging from the lens, it can be shown that this is the case if

$$\frac{\sin \theta_1}{\sin \theta_2} = \text{constant}$$

This is the so-called **sine condition**, and it can be shown that it is satisfied in the case of the object considered in Art. 46. The image formed of this object is therefore free from both spherical aberration and coma, and such a surface is called an **aplanatic surface**. Coma is reduced to a minimum by choosing a suitable shape for the surfaces of the lens, and this is investigated by ray tracing very much as was done for spherical aberration, and it turns out that the lenses which have the least spherical aberration have also the least coma.

50. ASTIGMATISM

Both spherical aberration and coma can be decreased by decreasing the radius of the stop sufficiently, but, even if they are completely eliminated,

the third term of our mathematical expression shows that another defect exists, which is serious for points in the image at an appreciable distance from the axis of the lens. It indicates that the emergent pencil passes through two focal lines perpendicular to each other and the principal ray, and so this defect of the image is known as astigmatism. The important thing to realise is that the radius of the circle of least confusion is proportional to the square of the distance of its centre from the axis of the lens, so that the defect is quite pronounced even though the aperture of the lens is small, the deviation from paraxial conditions being due to the large angle which the rays make with the axis on account of the large distance from it of the point from which they originate. The refracted pencil is astigmatic for the following reason. We showed in Art. 20 that a refracted pencil is astigmatic, when the focal length of the optical system through which it has passed is different in two mutually perpendicular planes. In an astigmatic lens this is due to one or both faces being cylindrical. In the

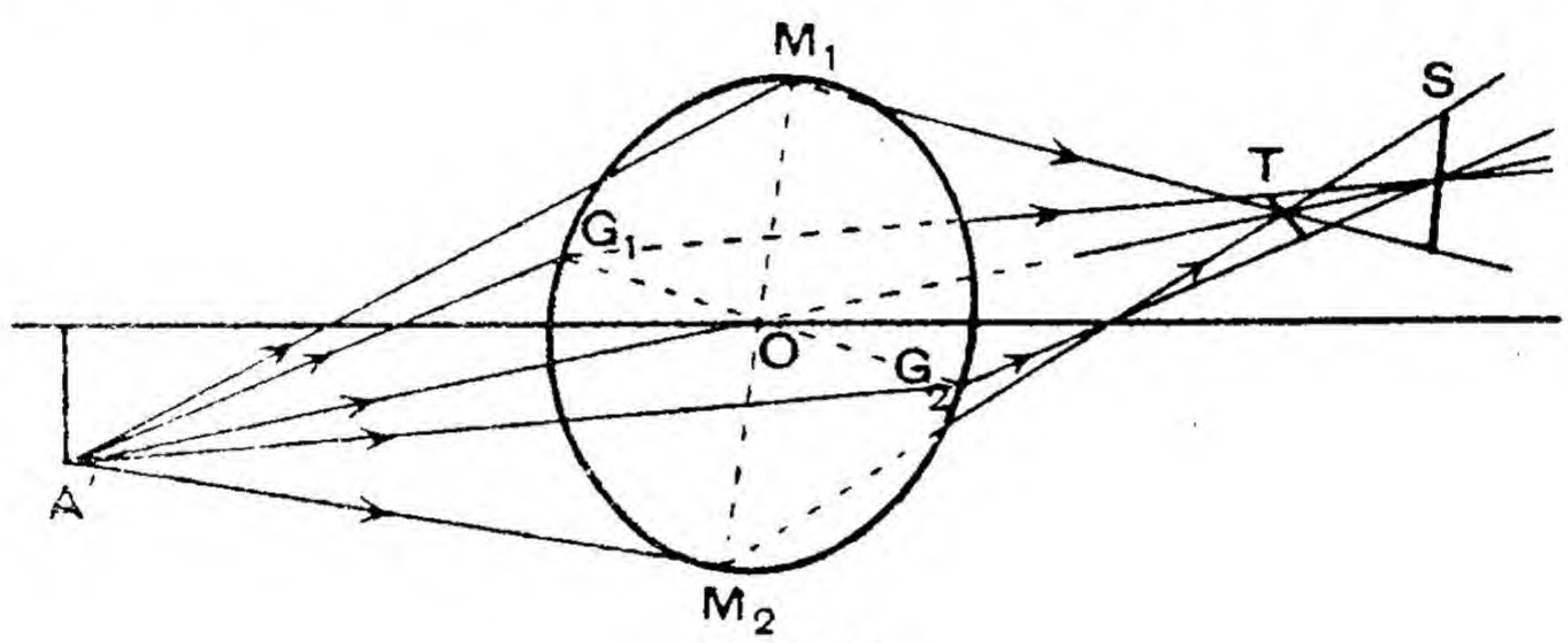


Fig. 86.

present case, while the point object is on the axis of the lens, the system is symmetrical, and so the focal length is the same in all planes through the axis of the lens and so the refracted pencil converges to a point image. But, as soon as the point object is off the axis of the lens, it is evident that symmetry has been lost and that the focal length of those rays in the plane containing the object and the axis, the **meridional plane**, is different from that of the rays in the plane through the object at right angles to this, called the **sagittal plane**. We shall show shortly from Fermat's principle that the focal length of rays in the meridional plane is less than that of those in the sagittal plane for a converging lens. The reader should be able to appreciate that the reason for this result is that the tilt of the lens is more pronounced for the rays in the meridional plane than for those in the sagittal plane. Assuming the truth of the above result, we can easily derive the nature of the refracted pencil. To consider the simple case in which the principal ray passes through the centre of the lens, let us imagine that the point object A' is vertically below the axis of the lens, which is horizontal, so that the meridional plane containing the extreme rays $A'M_1$ and $A'M_2$ is a vertical plane through the object and the axis of the lens, while the sagittal plane containing the rays $A'G_1$ and $A'G_2$ is one containing the object and a horizontal diameter

of the lens (Fig. 86). The rays in the meridional plane intersect at the tangential focus T and those in the sagittal plane at the sagittal focus S farther away from the lens. Consequently, since these two points do not coincide, the refracted pencil comes to a horizontal line focus, the tangential focal line, at T and to a vertical line focus, the sagittal focal line, at S, which is in the meridional plane and normal to the principal ray A′O. The circle of least confusion lies between these two line foci and is the best image which can be obtained of a point object off the axis of a lens. It follows that this defect spoils the definition of the edges of the image, which rapidly gets worse as we get farther from the axis, since the radius of the circle of least confusion is proportional to the square of its distance from the axis of the lens. It must be emphasised that astigmatism only assumes this simple form if both spherical aberration and coma are absent, and it should be added that coma itself is complicated by the presence of astigmatism.

It is interesting to compare this astigmatism with that produced by an astigmatic lens, for they are not quite the same. The reader will remember that an astigmatic lens produces a sharply focussed image of the wires of a gauze parallel to the axis of its cylindrical faces in one position and a sharply focussed image of the wires perpendicular to the axis of its cylindrical face in another position, but that it cannot focus both sets of wires at once. This is true even for those points of the image on the axis of the lens. The nature of the astigmatism of the oblique pencil passing through a lens with spherical faces is different, because the asymmetry is due, not to any asymmetry in the lens itself, but in the position of the object. In general, the tangential focal line is normal to the meridional plane, while the sagittal focal line lies in that plane, both lines being perpendicular to the principal ray. Therefore the direction of these lines alters as the position of the object is changed. To appreciate this, let us consider the nature of the image of a wheel with spokes placed with its centre on the axis of the lens. If a screen is placed normal to the axis of the lens at the tangential focal line for a point on the rim, the rim of the wheel will be sharply in focus, but the spokes will be out of focus. For, if the point of intersection of a spoke and the rim is considered as the object, it will form a line focus at the tangential focal line normal to the meridional plane, that is, perpendicular to the spoke, and so tangential to the rim of the wheel. This will be true of every point on the rim, and so the whole rim will be sharply in focus. This explains the reason for the name tangential focal line. If the screen is moved out to the sagittal focal line for a point on the rim, then the part of the spokes near to the rim will be in focus and the rim will go out of focus, since the sagittal focal line lies in the meridional plane and so lies parallel to the spokes. These lines point like arrows (sagitta) to the centre of the image, and this explains the name given to this focal line. We shall see below that the tangential and sagittal focal lengths decrease with increasing

obliquity, and so, as the screen is moved outwards, the portions of the spoke in focus move inwards towards the centre. The reader should work out for himself the nature of the image of an ordinary gauze formed by the lens. So we see that an astigmatic lens can focus sharply either of the sets of wires of a gauze, placed so that one set of wires is parallel to the axis of its cylindrical faces, and that the astigmatism persists right to the axis of the lens, whereas the astigmatism of an ordinary lens disappears for points on the axis of the lens, becomes worse the farther the point is from the axis, and consists in the ability to focus either a set of lines radiating from a point on the axis of the lens or a set of circles normal to those lines.

For the sake of those readers who are interested in the point, we shall now derive expressions for the tangential and sagittal focal lengths using Fermat's principle, but it is perhaps worth mentioning that it is not necessary to learn these proofs. They are merely inserted to show that the above statements do follow from the fundamental laws of rays of light. A parallel beam of light is incident on a converging lens at an angle θ to its axis, the principal ray striking the lens at G and pursuing

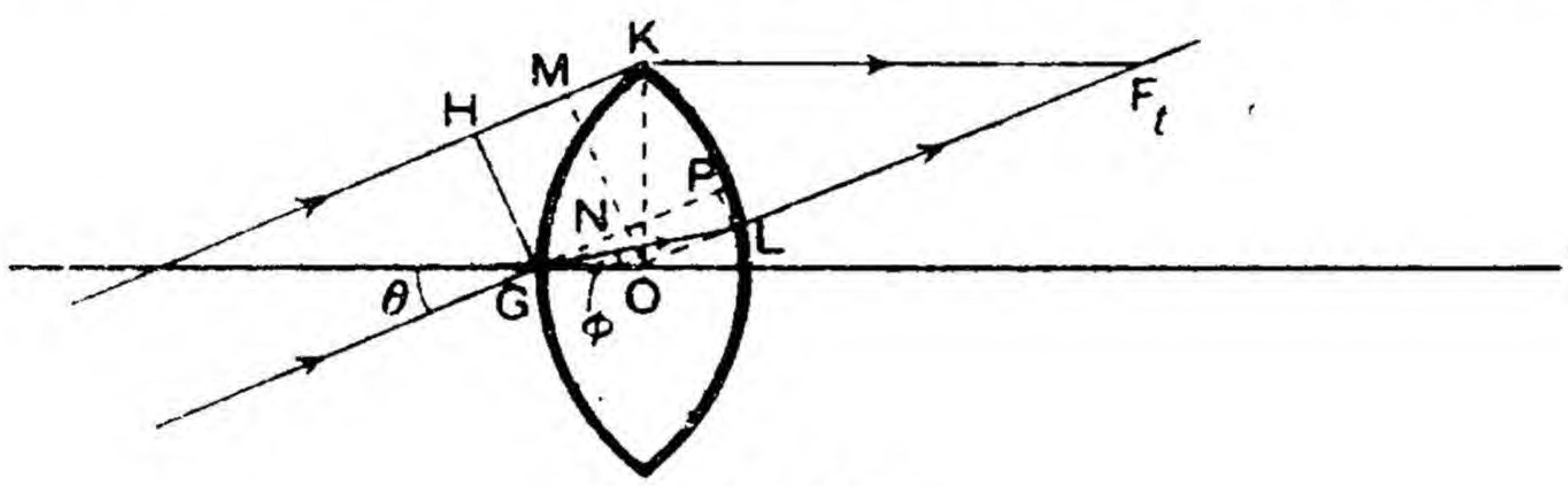


Fig. 87.

the path GLF_t, while the marginal ray follows the path HKF_t, so that F_t is the tangential focus of the beam, since the plane of the diagram is the meridional plane (Fig. 87). Let HG be normal to the beam and from the centre O of the lens draw OM normal to the direction of the incident beam to cut the marginal ray at M and the central ray produced at N. LP is a line from L normal to LF_t to cut GN produced at P and F_tL produced passes practically through O. If we assume that the lens is made of a material of refractive index n and is in air, we have from Fermat's principle

$$\frac{HK}{c}+\frac{KF_t}{c}=\frac{GL}{c_m}+\frac{LF_t}{c}$$

where c and c_m are the velocity of light in air and the lens respectively.

$$\therefore HK+KF_t=nGL+LF_t$$

$$\text{since} \frac{c}{c_m}=n$$

$$\therefore HM+MK+KF_t=n\,.\,GL+OF_t-OL$$

If t=the thickness of the lens

h=OK, the radius of the lens

f_t=OF$_t$, the tangential focal length

$$\therefore\ \text{GN}+h\sin\theta+\{h^2\cos^2\theta+(f_t-h\sin\theta)^2\}^{\frac{1}{2}}=\frac{nt}{\cos\phi}+f_t-\text{NP}$$

$$\therefore\ h\sin\theta+(f_t-h\sin\theta)\left\{1+\frac{h^2\cos^2\theta}{(f_t-h\sin\theta)^2}\right\}^{\frac{1}{2}}=\frac{nt}{\cos\phi}+f_t-\text{GP}$$

If h is small, we may expand the term to the power $\frac{1}{2}$ by the binomial theorem and neglect terms in h^4 and higher powers ; so we have

$$h\sin\theta+f_t-h\sin\theta+\frac{h^2\cos^2\theta}{2(f_t-h\sin\theta)}=\frac{nt}{\cos\phi}+f_t-\frac{t}{\cos\phi}\cos(\theta-\phi)$$

$$\therefore\ \frac{1}{f_t-h\sin\theta}=\frac{2t}{h^2\cos^2\theta}\left\{\frac{n}{\cos\phi}-\frac{\delta\cos(\theta-\phi)}{\cos\phi}\right\}$$

But $$\frac{n}{\cos\phi}-\frac{\cos(\theta-\phi)}{\cos\phi}=\frac{n}{\cos\phi}-\frac{\cos\theta\cos\phi+\sin\theta\sin\phi}{\cos\phi}$$

$$=\frac{h}{\cos\phi}-\cos\theta+\frac{n\sin^2\phi}{\cos\phi}=n\cos\phi-\cos\theta,\ \text{since } n=\frac{\sin\theta}{\sin\phi}$$

$$\therefore\ \frac{1}{f_t-h\sin\theta}=\frac{2t}{h^2\cos^2\theta}(n\cos\phi-\cos\theta)$$

But from Art. 19

$$\frac{2t}{h^2}=\left(\frac{1}{c_1}+\frac{1}{c_2}\right)=\frac{1}{f(n-1)}$$

using the usual notation.

$$\therefore\ \frac{1}{f_t-h\sin\theta}=\frac{1}{f}\cdot\frac{n\cos\phi-\cos\theta}{(n-1)\cos^2\theta}$$

The marginal ray at the opposite end of the diameter of the lens through K meets the central ray at a distance given by

$$\frac{1}{f_t+h\sin\theta}=\frac{1}{f}\cdot\frac{n\cos\phi-\cos\theta}{(n-1)\cos^2\theta}$$

If h is small, these two points coincide and we have for f_t the expression

$$\frac{1}{f_t}=\frac{1}{f}\cdot\frac{n\cos\phi-\cos\theta}{(n-1)\cos^2\theta}$$

To see how this compares with f, it is convenient to express all the angles in terms of $\sin\theta$.

$$\therefore\ \frac{1}{f_t}=\frac{1}{f}\cdot\frac{n\left(1-\frac{\sin^2\theta}{n^2}\right)^{\frac{1}{2}}-(1-\sin^2\theta)^{\frac{1}{2}}}{(n-1)(1-\sin^2\theta)}$$

Expanding by the binomial theorem and neglecting terms higher than $\sin^2\theta$, we have

$$\frac{1}{f_t}=\frac{1}{f}\cdot\frac{(n-1)\left(1+\frac{\sin^2\theta}{2n}\right)}{(n-1)(1-\sin^2\theta)}$$

$$\therefore \frac{1}{f_t}=\frac{1}{f}\left(1+\frac{2n+1}{2n}\sin^2\theta\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

So we see that the tangential focal length measured along the principal ray is less than the paraxial focal length for a converging lens.

The calculation of the sagittal focal length f_s follows the same lines; the path of the principal ray through the centre of the lens is the same as above, but the two marginal rays are in phase with it, when they reach the lens. So we have

$$\frac{1}{f_s}=\frac{2t}{h^2}\left\{\frac{n}{\cos\phi}-\frac{\cos(\theta-\phi)}{\cos\phi}\right\}$$

$$\therefore \frac{1}{f_s}=\frac{1}{f}\cdot\frac{n\cos\phi-\cos\theta}{(n-1)}$$

$$\therefore \frac{1}{f_s}=\frac{1}{f}\cdot\left(1+\frac{\sin^2\theta}{2n}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

Again, the sagittal focal length measured along the principal ray is less than the paraxial focal length but greater than the tangential focal length. From equations (37) and (38) we have

$$\frac{1}{f_t}-\frac{1}{f_s}=\frac{1}{f}\cdot\sin^2\theta$$

$$\therefore f_s-f_t \propto \sin^2\theta.$$

This shows that the two line foci get rapidly farther apart, as the obliquity of the pencil increases, and hence the radius of the circle of least confusion will rapidly increase, as the distance of the point object from the axis increases. So this defect of astigmatism means that the definition falls off rapidly towards the edges of an image of a large object. The correction of astigmatism is closely related to that of another defect, curvature of the field, and it will therefore be considered along with the correction of this defect.

51. CURVATURE OF THE FIELD

This defect does not occur as a term in the mathematical expression, because that expression refers to the radius of the circle of least confusion formed round the paraxial focus. It is now necessary to inquire where the most sharply focussed image of objects a small distance from the axis of the lens will be situated relative to the paraxial focus. If A is a point object on the axis of a convex lens, which is stopped down so that only narrow

central pencils can pass through it, the position of its paraxial image B can be found from the usual relation

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

If the line AOB is rotated through a finite angle about O in the plane of the paper, then A describes the arc PAQ and B the arc RBS (Fig. 88). If we assume that the focal length measured along the principal ray is independent of its angle with the axis, it is evident that RBS is the image of PAQ. For consider the point P; the principal ray of the pencil from P going through the lens is POS, and as the focal length along this ray is the same as that of the lens and PO=AO, then the image of R is the

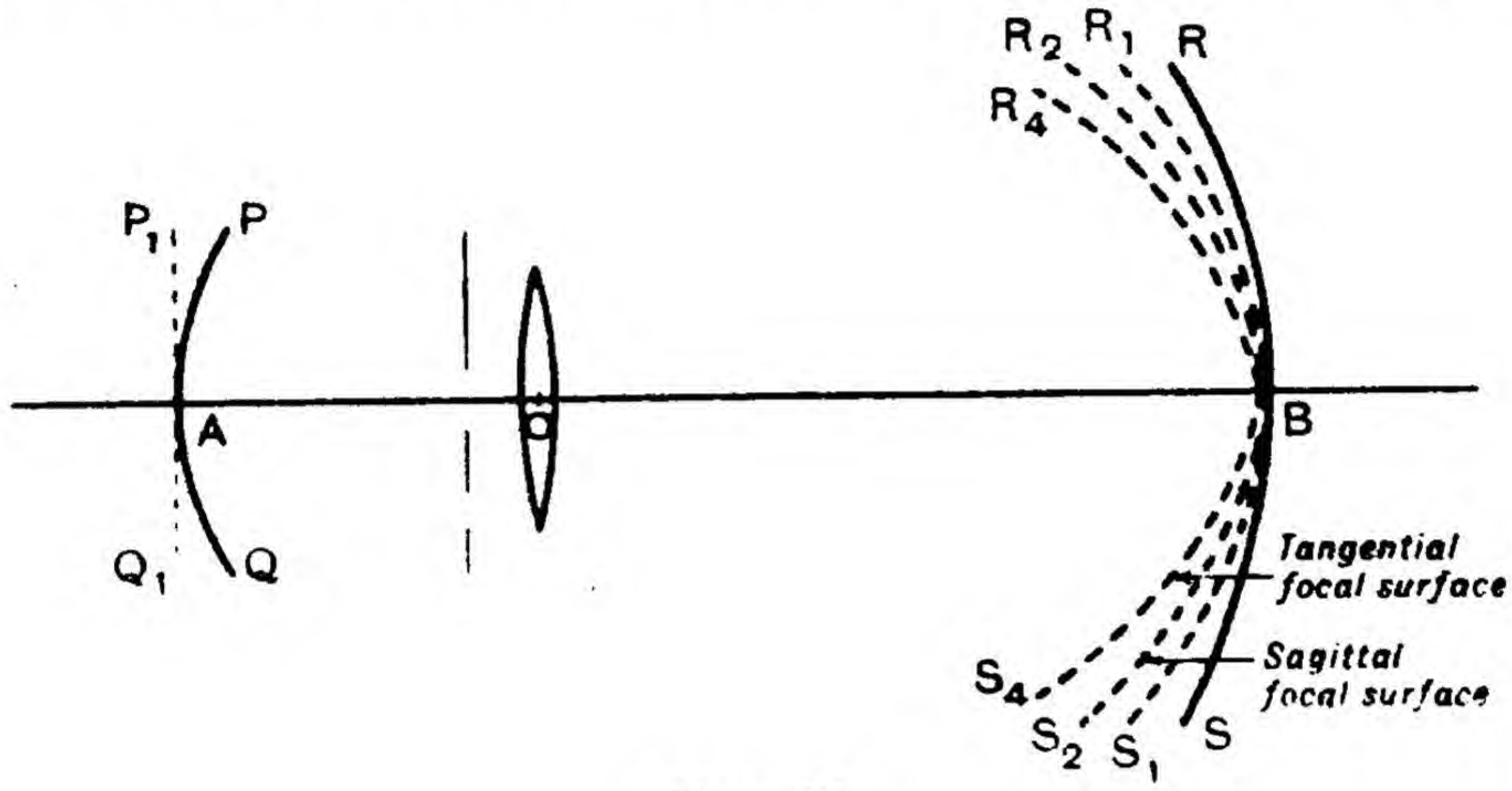


Fig. 88.

same distance from O along POS as OB, and so it is at S. If the diagram is now rotated about the axis of the lens, PAQ and RBS become portions of the surfaces of spheres with a common centre at O, and it follows that the image of a portion of the surface of a sphere whose centre is at O will also be a portion of the surface of another sphere with centre at O. As a rule objects of finite size are flat and we can deduce what will happen in this case, if we imagine the surface of the sphere to be flattened until it becomes a plane P_1AQ_1 perpendicular to the axis of the lens. Again consider the point P. Owing to the flattening of the spherical object surface it has moved back along the line SOP to P_1. Now it can easily be shown from the usual thin lens formula that, if a point object is moved in any direction along the axis or principal ray, the image moves in the same direction. (See Example at the end of Chapter 2.) So, if P moves to P_1 to transform the spherical object surface into a plane object, the image point S moves inwards to S_1 to form an image surface R_1BS_1, which will have a greater curvature than before. So the paraxial image of a plane object of finite area is a curved surface concave towards the object in the case of a converging lens. We have finally to take account of the fact, which was shown in the previous article, that both the tangential and sagittal focal lengths for oblique rays are less than the paraxial focal length. It is evident that this will cause both

the tangential and sagittal focal surfaces to be of even greater curvature than the paraxial focal surface as shown in Fig. 88. The most sharply defined image will lie at the circle of least confusion which lies on a curved surface situated between the tangential and sagittal surfaces (Fig. 89). Since it is frequently necessary to throw an image of a flat object on to a flat screen, this defect of curvature of the field will still

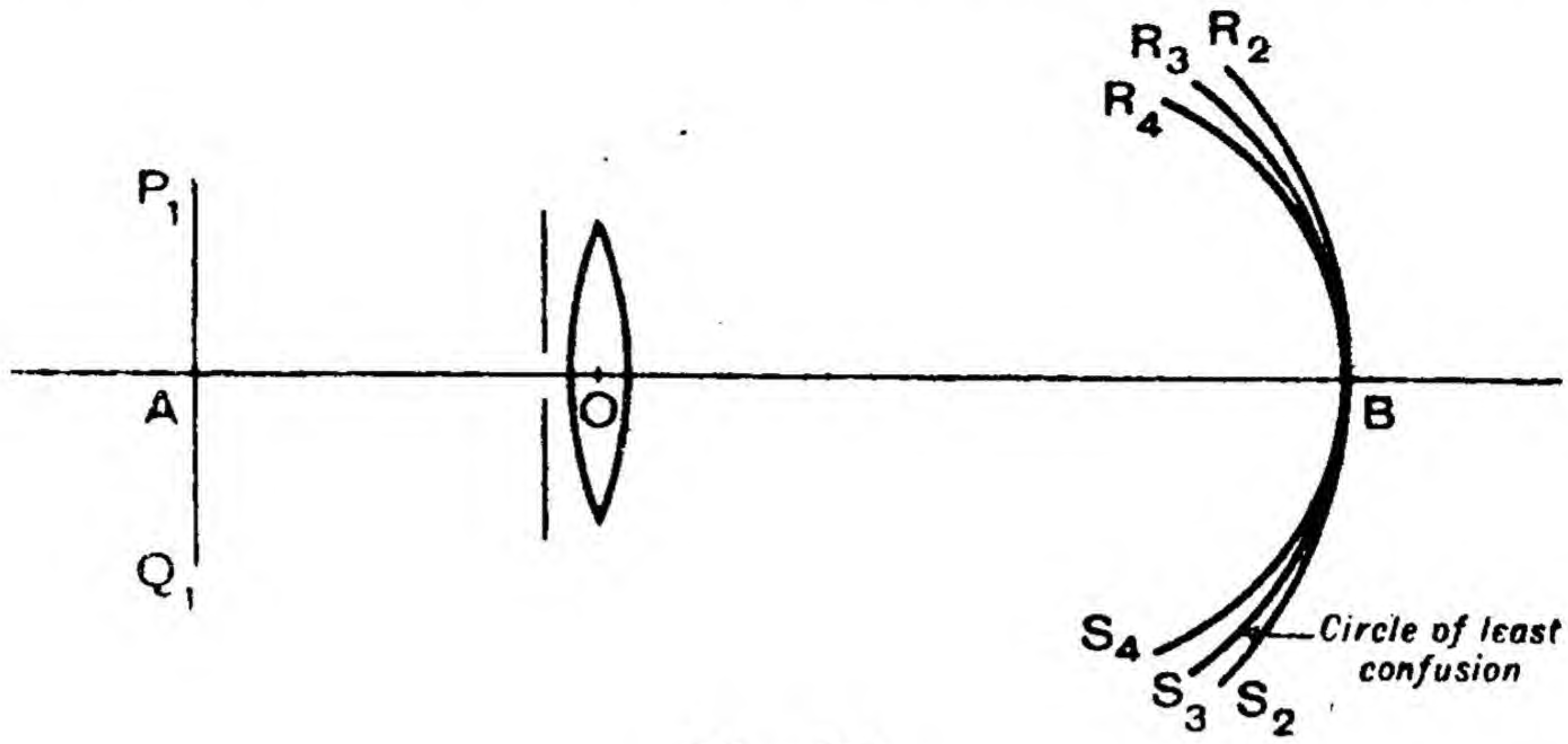


Fig. 89.

further increase the loss of definition at the edges of the field due to astigmatism.

The reader will see at once from Fig. 90 that the curvature of the field produced by a diverging lens forming a virtual image of a real object is in the opposite direction to that produced by a converging lens. The transformation of PAQ into the flat object P_1AQ_1 will tend to make the image RBS more nearly plane, but this is counteracted by the fact that the focal length of oblique rays measured along the principal ray

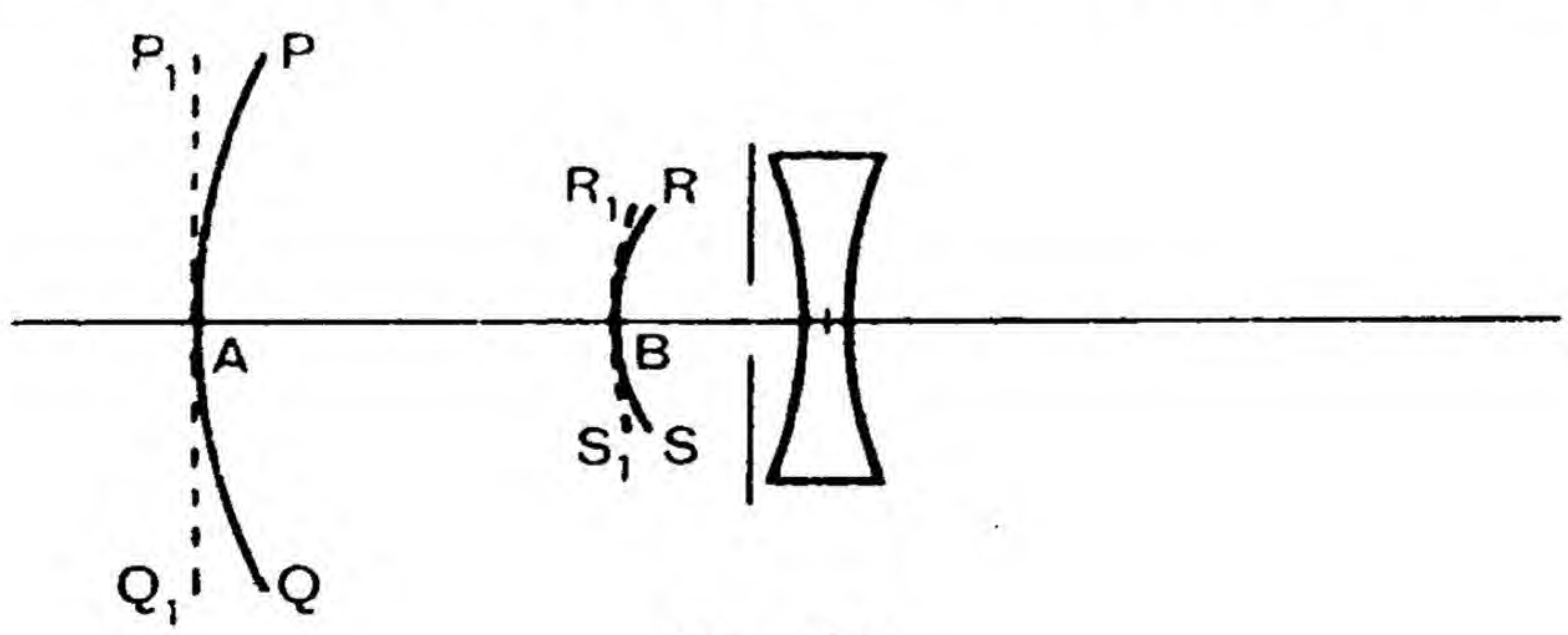


Fig. 90

is less than that of paraxial rays. This will cause the image of P to move towards the centre of the lens and so the final image will be in a position such as R_1BS_1. This suggests that the field may be flattened by a suitable choice of converging and diverging lenses and the following equation, known as the **Petzval condition,** must be satisfied to accomplish this :

$$n_1f_1+n_2f_2=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

where n_1 and f_1 are the refractive index and focal length of the first lens and the corresponding letters with the subscript 2 represent the same

quantities for the second lens. This condition is valid for any distance apart of the lenses and, if it is satisfied, astigmatism is also corrected.

It should be mentioned that curvature of the field is largely due to the change in focal length measured along the principal ray with obliquity, since all the above treatment only applies to objects so close to the axis that the curvature of the surface RBS (Fig. 88) will be very small. The above results are also true if the stop is separated from the lens by a finite distance.

52. DISTORTION

The first three defects spoil the definition of the image of a point object off the axis, but this last one, which arises from the fourth term in our mathematical expression, refers to its position relative to the paraxial focus. It resembles curvature of the field in this respect, although we saw that curvature of the field did affect the definition of the edges of an image if it was cast on a flat screen. But, whereas curvature of the field refers to the displacement of the actual image from the paraxial focus in a direction along the principal ray, distortion refers to its displacement along the line from the paraxial image perpendicular to the axis

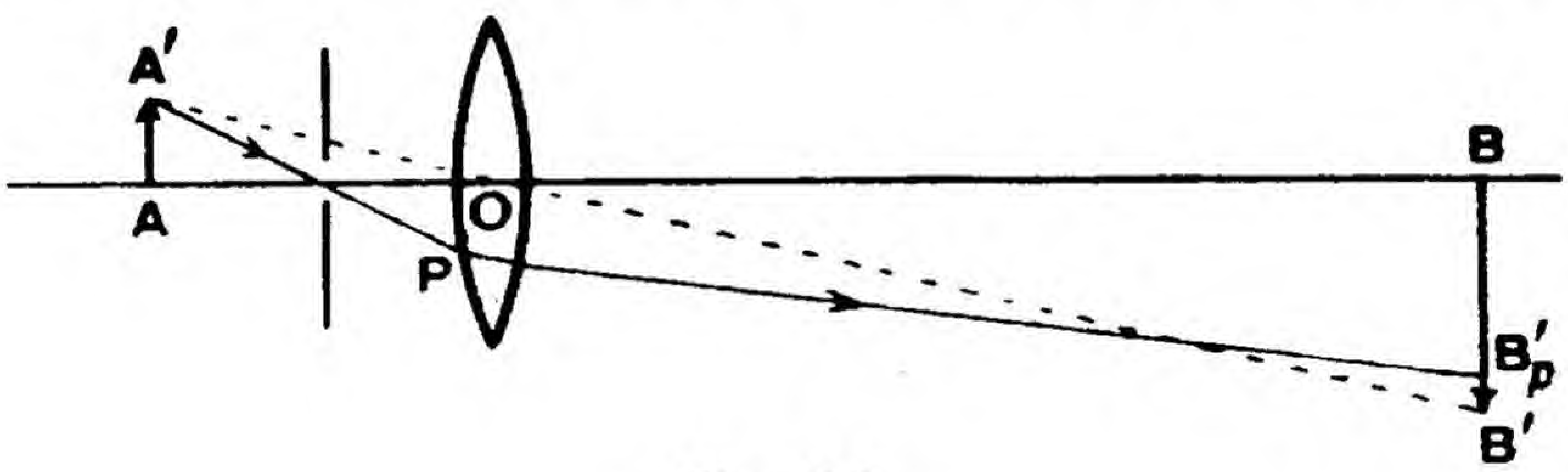


Fig. 91.

of the lens. To understand how it arises it is necessary to imagine that all the previous defects have been corrected out, so producing a final point image somewhere on the principal ray. It is necessary to trace the course of this ray through the lens. In Fig. 91 B is the paraxial image of a point object A on the axis of the lens and B′ is the paraxial image of the point object A′ and so lies on the straight line A′OB′, the auxiliary axis of the lens. The principal ray A′P passes through the lens at a finite distance from the auxiliary axis A′OB′, so that it will cross that axis nearer to the lens than the rays very close to the axis and intersect the plane through B perpendicular to the axis of the lens, on which the image is being cast, at B_p'. Hence the final point image of A′ is formed at B_p' nearer to the axis of the lens than the paraxial focus and the amount of the distortion is measured by the distance $B'B_p'$ and is reckoned positive in the case shown. The full mathematical theory shows that the distortion is proportional to $BB_p'^3$. It also increases as the stop is moved further from the lens, since the principal ray crosses the lens further from the auxiliary axis, and for the same reason it is zero when the stop coincides with the lens. It is easy to see that the effect of this defect is to distort the image, and in the case of positive distortion the outer parts are closer to the axis than in the

paraxial image. Let us imagine that the object is a square whose centre is at A and the mid-point of one side is at A′, then the paraxial image has its centre at B and the mid-point of one of its sides at B′ and is shown in Fig. 92 in dotted lines. The actual image of A′ is at B_p′, and that of the whole square is shown in continuous lines, the distortion of the corners being greater than that of a point such as B′, since they are further from the axis. This kind of distortion is often called barrel distortion, since the image of the square looks like a barrel; we see that it occurs, when the stop is between the lens and the object, and that the central parts of the image are magnified relative to the outer parts. (Plate III, Fig. 2.)

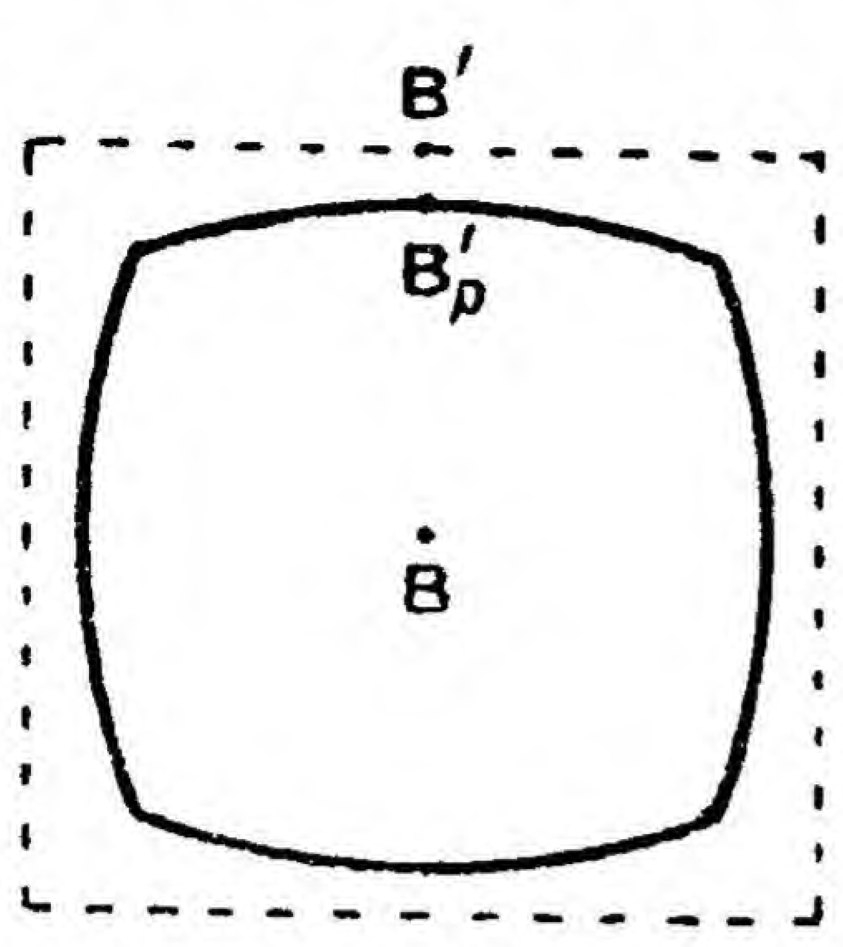


Fig. 92.

If the stop is placed between the lens and the image, an examination of Fig. 93 shows that the principal ray crosses the plane through B perpendicular to the axis of the lens further from the axis than the paraxial focus.

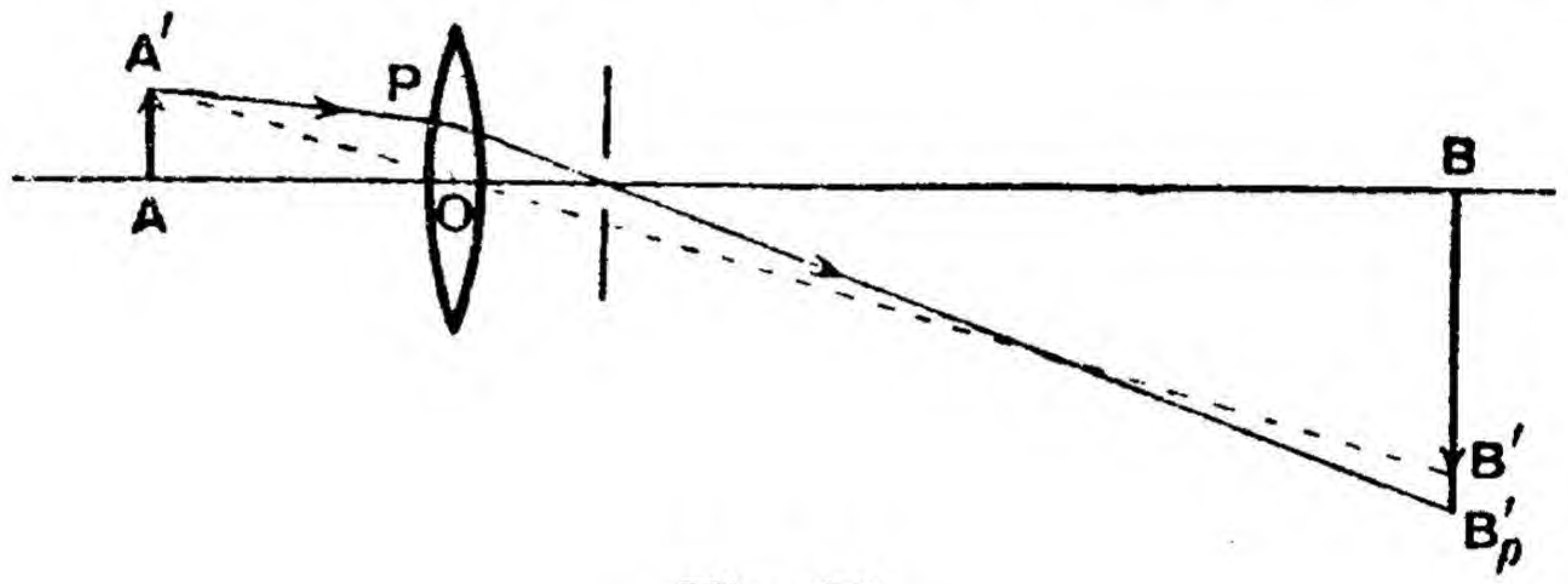


Fig. 93.

This is the case of negative distortion and the actual image of a square is shown in Fig. 94, which accounts for this type being called pincushion distortion. The reader will notice that here the outer parts of the image are magnified relative to the central parts. It is essential to realise that this defect is a defect of the principal ray and that it persists, even if the width of the stop is infinitely small; it must also be emphasised that it is not due to spherical aberration in the sense in which it has been defined in this chapter, for it persists after that defect has been removed. So it is not correct, as is sometimes done, to attribute it to the fact that the focal length measured along the axis of the lens is less for marginal than for paraxial rays, for these focal lengths are the same after spherical aberration has been corrected. What we have learned about distortion suggests a way of correcting it by making a lens system of two separated components with a stop between them, so that the pincushion distortion due to the first is corrected by the barrel distortion due to the second. (Plate III, Fig. 3.)

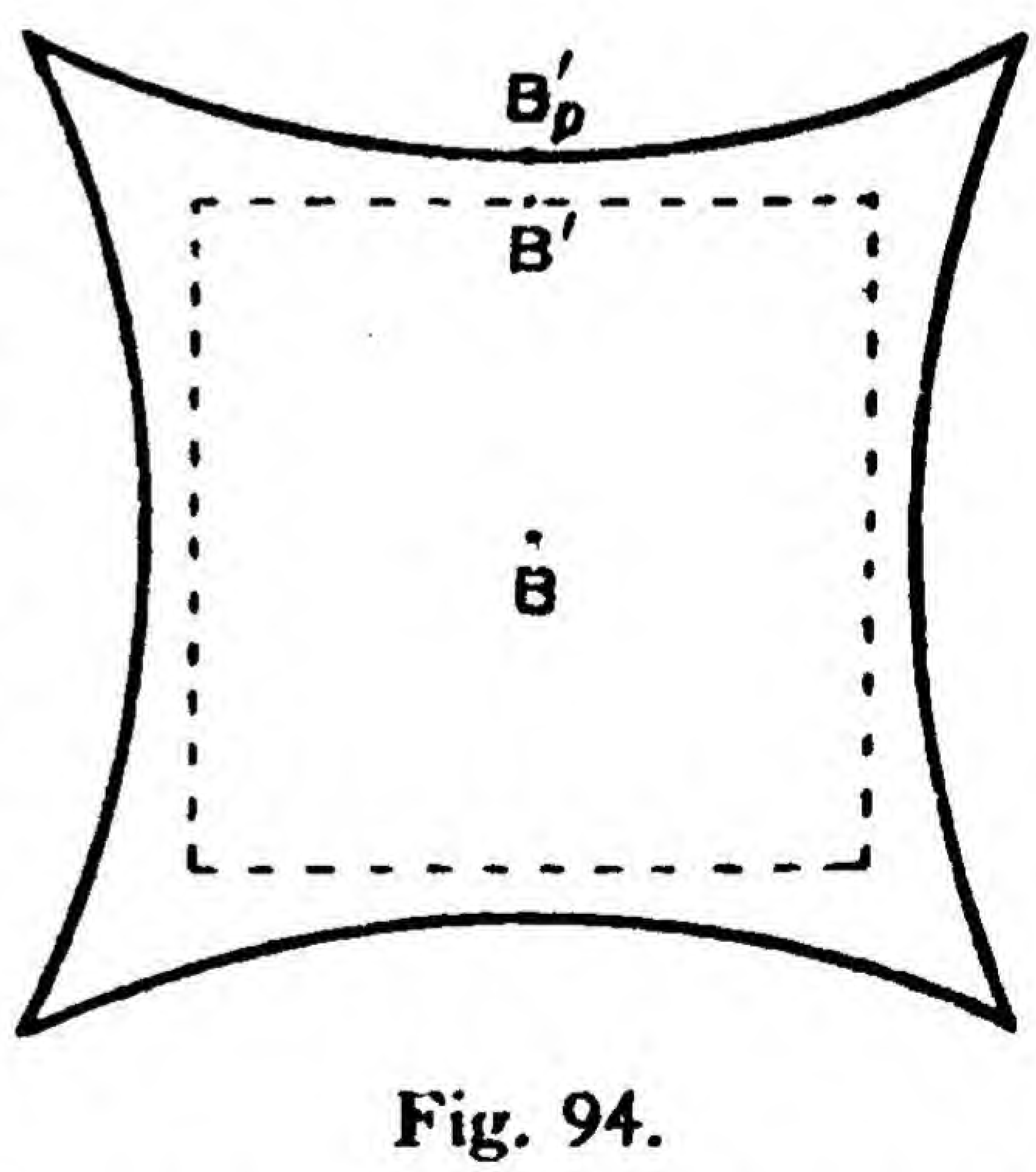


Fig. 94.

PLATE III

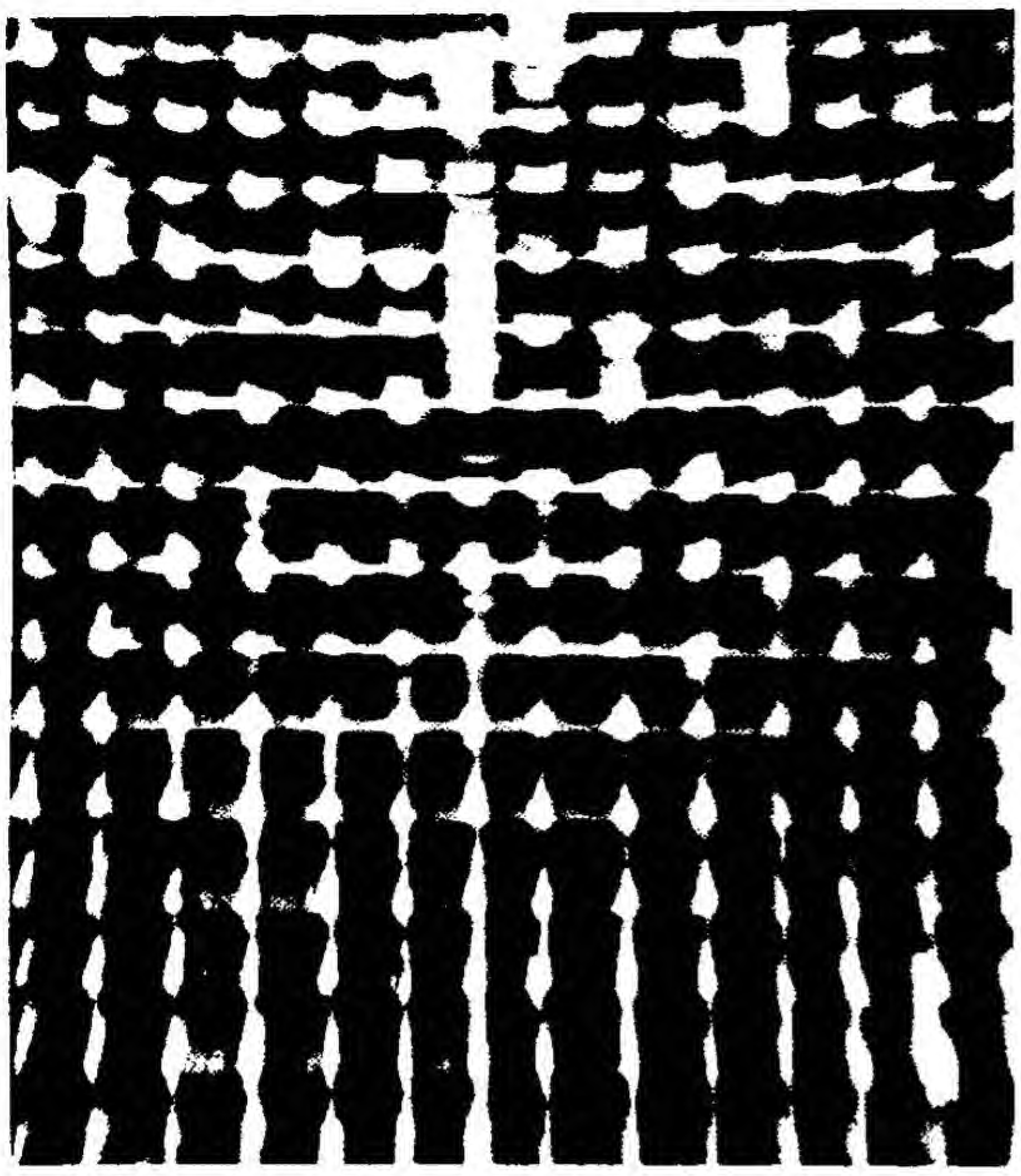


Fig. 1. The image of a gauze formed by rays passing obliquely through the centre of the lens showing astigmatism.

(J. W. Mitchell)

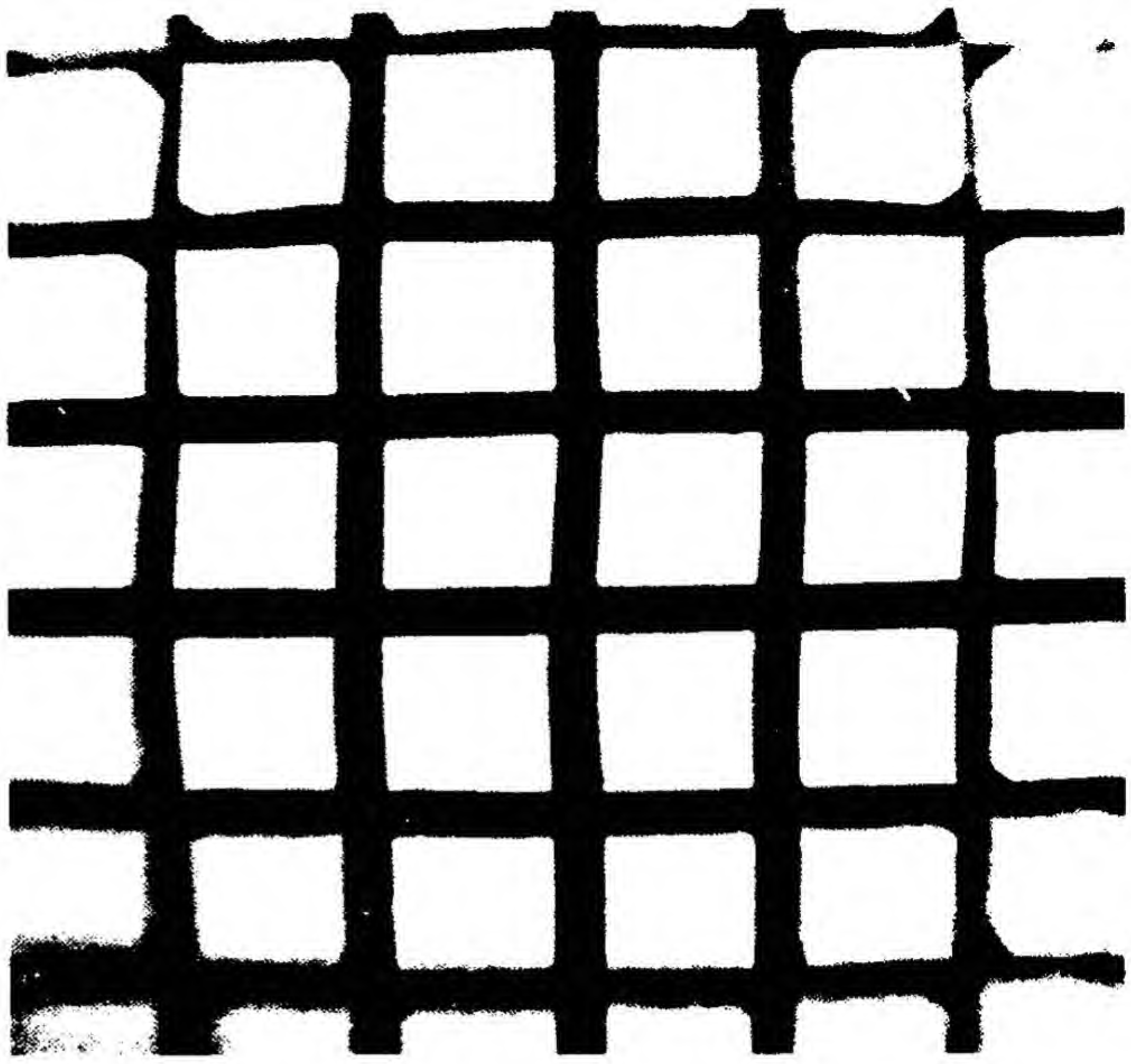


Fig. 2. Barrel Distortion.

(J. W. Mitchell)

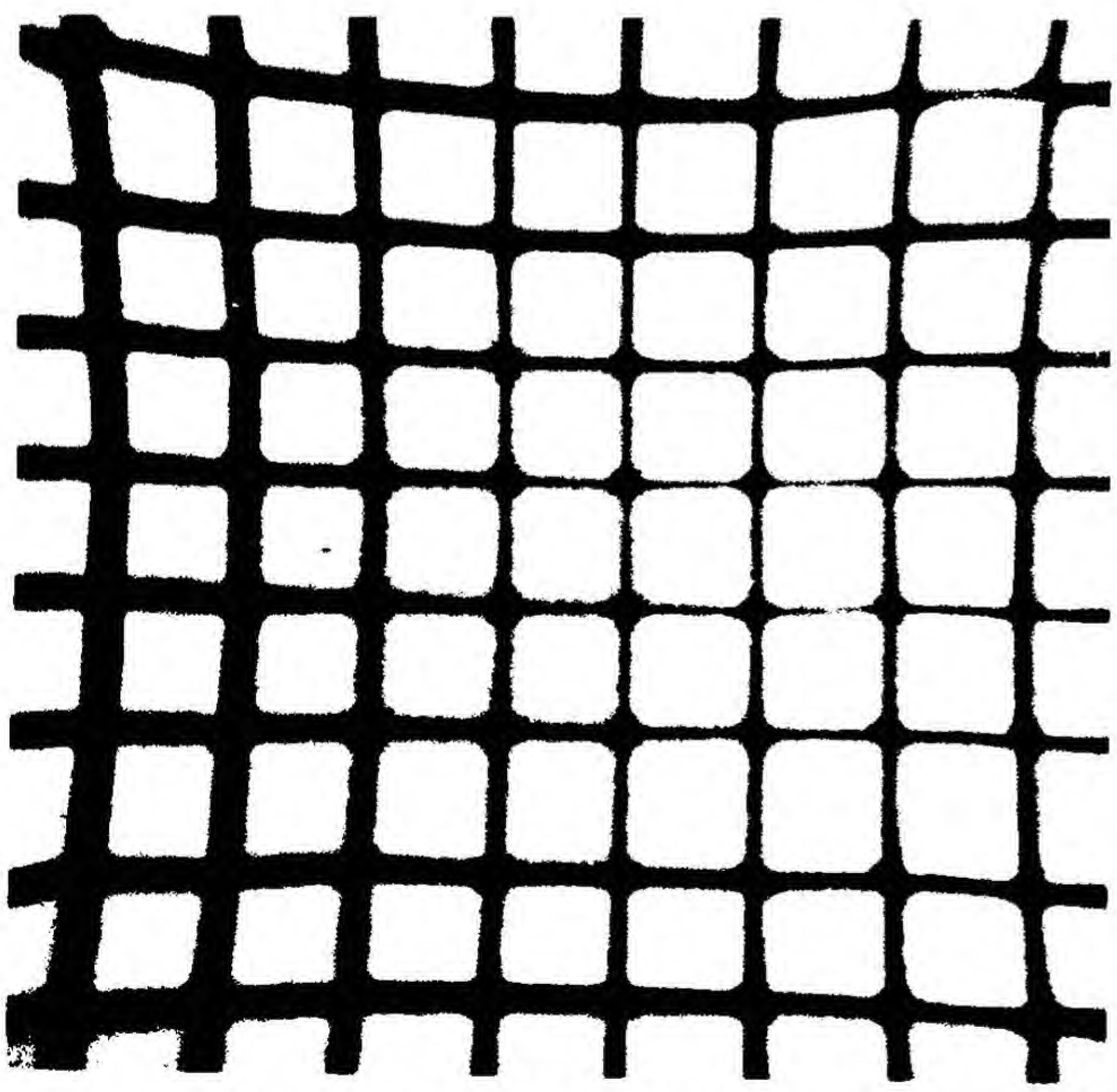


Fig. 3. Pincushion Distortion.

(J. W. Mitchell)

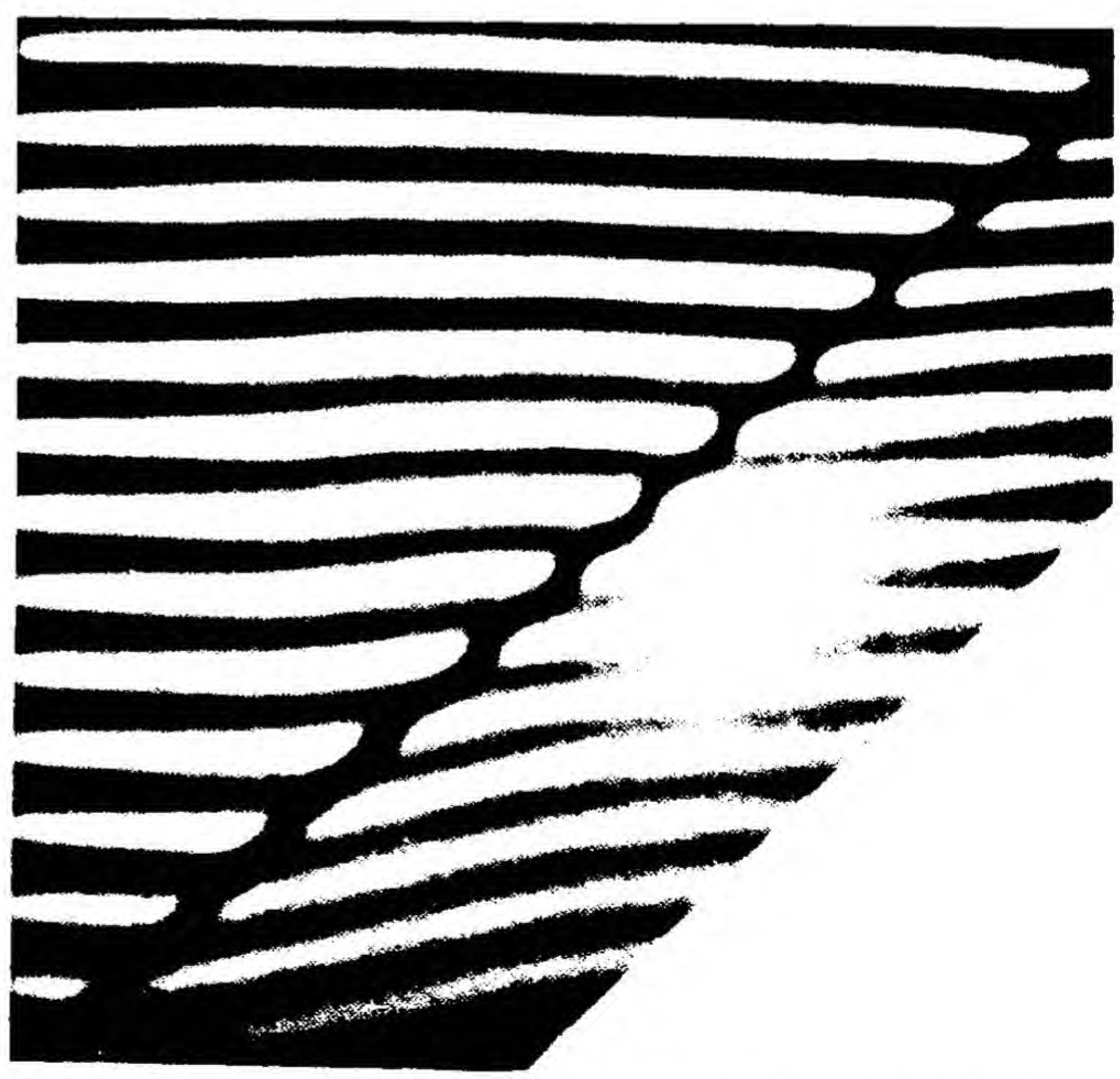


Fig. 4. Refraction of plane waves in passing from deep water in the upper left-hand part of the photograph to shallow water in the lower right-hand part of the photograph.

(J. W. [illegible])

53. CONCLUSION

We may sum up this important account of the defects of the image formed by a single lens in this way. Rays from a point object on the axis of a lens passing through the same zone of the lens come to a point focus on the axis, but the greater the radius of the zone, the nearer the focus is to the lens. This is spherical aberration and is measured by the distance from the paraxial focus to the focus of the given zone and is proportional to the square of the radius of the zone. This result follows simply from Snell's law and the spherical shape of the surfaces of the lens. It reduces the definition at the centre of the image of a large object. The same thing is true of the results obtained for points off the axis of the lens, where the case is complicated by the asymmetrical position of the object. The actual path of the various rays in the emergent beam is calculated ultimately from Snell's law and the form of the expression leads to the properties of the beam being naturally split into five parts. The first three affect the definition of the best approach to a point image formed by the beam ; these are spherical aberration, coma, and astigmatism. The first is just the same defect as occurred with points on the axis, the other two arise from the asymmetrical position of the object, coma being an asymmetrical deformation of the image and astigmatism the presence of two mutually perpendicular focal lines at right angles to the principal ray. The last two properties of the beam refer to the position of the image, curvature of the field giving its distance from the paraxial image in the direction of the principal ray and distortion its distance at right angles to that ray. It cannot be too strongly emphasised that these defects only occur in the simple form in which they have been described, provided that the previous ones have been eliminated and, in the case of coma, if astigmatism also has been eliminated. Since they all follow from Snell's law, the spherical shape of the lens surfaces, and the position of the object, it follows that, in practice, they are not independent. Consequently, if a lens is uncorrected in any way, they are all present together and the reader must try to get some idea of the resulting emergent pencil. There will be two constrictions in the beam at the tangential and sagittal foci, but they will have a finite width ; there will be an asymmetrical deformation of the image corresponding to coma, but the comatic circles have degenerated into loops due to the astigmatism and so on. It also follows from this interdependence of the aberrations that the change of any one aberration due to a shift in the position of the stop depends partly on the amount of the preceding aberrations present. It is therefore sometimes advisable not to correct out an aberration completely, in order that it may help to correct some later one. This is called the **balancing of aberrations** and is used in lens designing. We shall hope to show how the principles established here are used in designing the components of the various optical instruments as we come across them, but we may

notice an important general principle in conclusion. In designing a thin lens to be as free of spherical aberration as possible, the usual paraxial relation is used to settle what pairs of radii of curvature are permissable for the faces and then the pair giving the least spherical aberration is found by ray tracing through the various permitted possibilities. So we see that the function of the paraxial relation is to limit the number of possibilities to which ray tracing is applied and so to shorten the labour of this trial and error method. Precisely the same thing is true in designing lenses which are free from the other aberrations, the full mathematical theory being used to limit the number of possibilities to which ray tracing is applied.

EXAMPLES ON CHAPTER VI

1. What do you understand by spherical aberration?
The surface of greater curvature of a concavo-convex lens has a radius of curvature a and the centre of curvature of the second surface is distant $(1+1/\mu)a$ from the pole of the first. Find the position of the image formed by a point object placed at the centre of the second curvature and show that no approximation is required concerning the aperture of the lens. *(Oxford Schol.)*

2. State clearly and fully what is meant by spherical aberration. Discuss the various ways by which it is reduced to a minimum, bringing out the physical principles involved.

3. In what way does spherical aberration spoil an image? Can it ever be completely eliminated? If so, describe fully how it can be done, proving that a perfect point image is formed of a point object.

4. Lens aberrations are due to the violation of paraxial conditions. What are the two main ways in which such violation can occur? Name the defects produced in each case and indicate briefly their effect on the image of a point object formed by a single lens.

5. Why do aberrations occur in an oblique narrow pencil passing through a lens, when there would be no aberration if the same pencil passed through the centre of the lens in a direction parallel to the axis? Discuss briefly the five aberrations produced and indicate their effect on the image of an object of finite size placed normal to the axis of the lens.

6. What are astigmatism, curvature of the field, and distortion? What is the cause of them and how may they be reduced to a minimum?

7. Describe the usual defects in the image of an object formed by a lens. What means are employed to correct them? *(Camb. Schol.)*

8. Explain why pincushion distortion is produced if a virtual image of an object is formed by a convex lens and looked at by the eye at some distance from the lens, while barrel distortion is produced if the lens forms a real inverted image which is viewed by the eye.

Chapter VII

THE EYE AND THE CAMERA

54. STRUCTURE OF THE EYE

We have now established the main theorems of geometrical optics, having derived the concept of the thin lens in the course of our analysis and having seen that it represents very closely the behaviour of actual lenses under paraxial conditions. We have also seen in a general way what modifications our simple paraxial theorems undergo, when rays incident at larger angles are considered, and it is now time to apply these theorems to the various optical instruments, in order that we may see both how they achieve their purpose and how their design may be improved. It is natural to begin with the human eye and the camera, since their optical principles, at any rate, are fairly simple and resemble each other closely.

The eye is an organ of globular shape, about an inch in diameter, which is held in position in its socket by a set of muscles, which can also turn it about so as to make its optic axis point in various directions. It is enclosed in a hard opaque coat, S, the sclerotic (Fig. 95), which becomes a transparent window, the cornea C, at the front of the eye. The **crystalline lens** L is held behind this window by the suspensory ligaments attached to the ciliary muscles M, which can alter the power of the lens by contracting and so altering the curvature of its faces. Inside the sclerotic

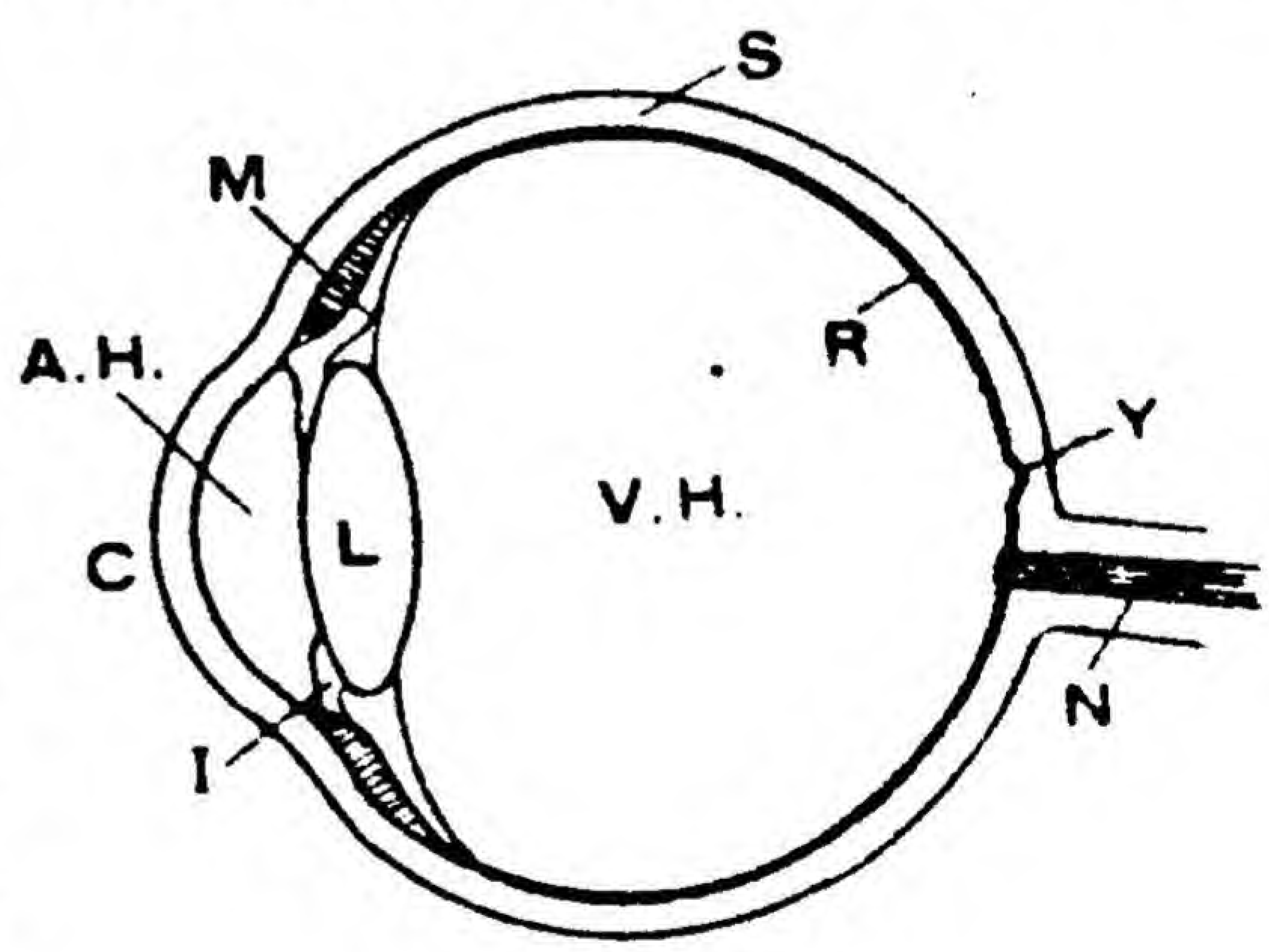


Fig. 95.

there is a coating called the **retina** R, which is sensitive to light, since the whole of its surface is covered with the nerve endings of the individual fibres of the optic nerve N, which enters the eye at the back and branches out in all directions so as to send fibres to all parts of the retina. These nerve endings consist of rods and cones and, when light falls on either a rod or a cone, an impulse is sent up the corresponding nerve fibre and a message is conveyed to the brain and the person receives the sensation of sight. The cones are larger than the rods, being from 0·0011 to 0·0054 mm. in diameter, but there are 18,000,000 rods to 3,000,000 cones. There is a sensitive region in the retina, a depression

at Y, called the **yellow spot**. It is about 2 mm. across and in the middle of it is a smaller area, the fovea, about 0·25 mm. in diameter, which is the most sensitive place of all on the retina; it consists entirely of cones. The space between the cornea and the lens is filled with the aqueous humour, AH, a salt solution, while that between the lens and the retina is filled with a kind of watery jelly, called the vitreous humour, VH. Lastly there is a coloured diaphragm with an aperture of variable diameter in front of the lens. It is called the **iris** I and it is this part of the eye which is referred to, when the colour of a person's eye is mentioned. The black hole in the middle of the iris is called the **pupil** and it looks black to an observer because, when he looks at it, he is looking into the person's eye, from which very little light emerges since the retina absorbs nearly all the light which falls on it. The diameter of the pupil is automatically controlled by the amount of light falling on the eye, and it varies from 2 mm. to 8 mm. according to the brightness of the field at which the eye is looking in accordance with the relation shown in Fig. 96.

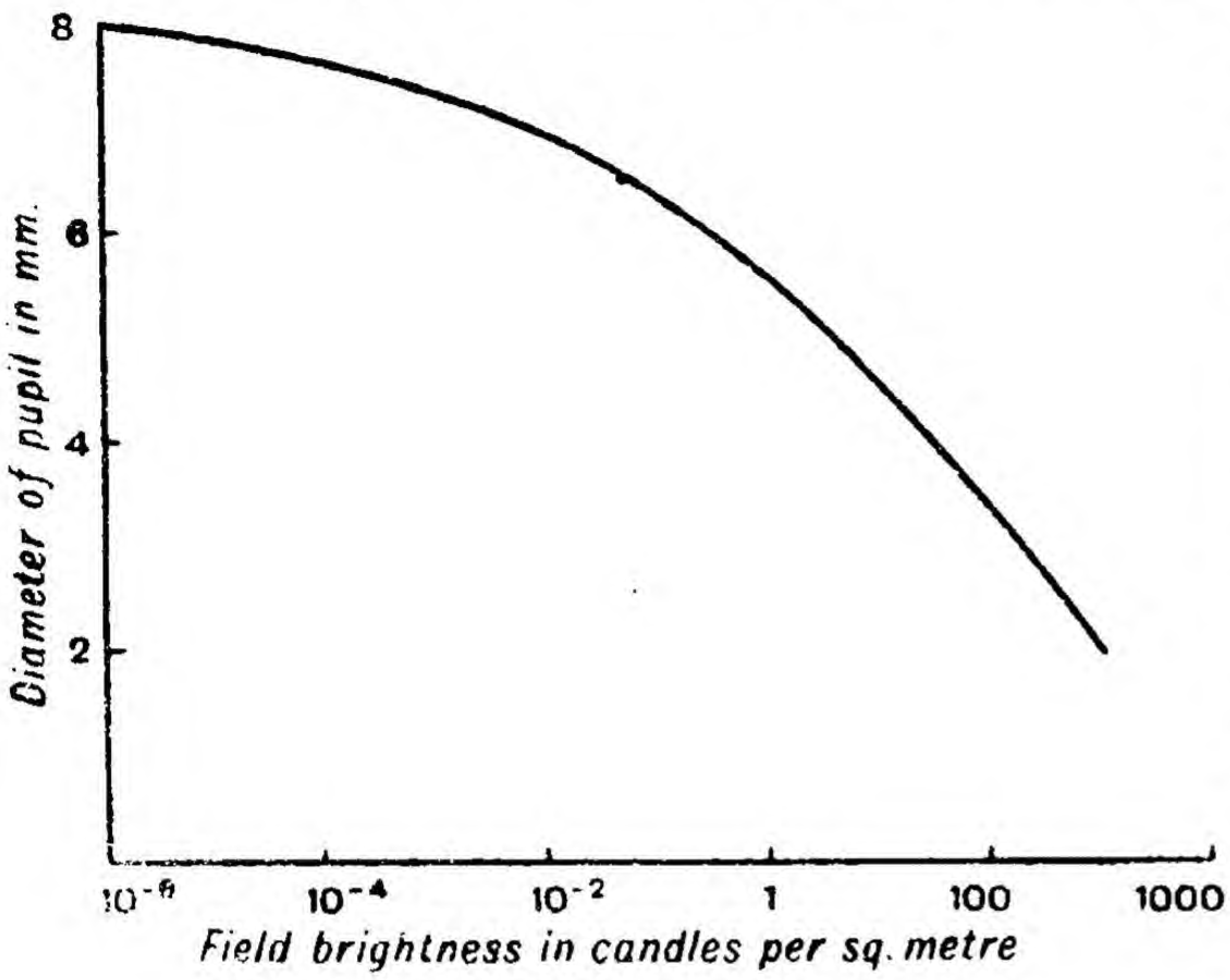


Fig. 96.

55. NORMAL VISION AND ACCOMMODATION

It is desirable to be acquainted with the order of magnitude of the various parts of the eye in order to understand how objects are seen, and a table is given of these magnitudes when the eye is looking at a point at infinity and at a point 152·5 mm. away (Table 8). All the quantities are measured directly except the position of the two principal foci. The refractive index of the crystalline lens is a mean value, since it is greater for the outer than the inner layers. It is clear from these numbers that *most of the refraction takes place at the cornea*, as there is so little difference in refractive index between the crystalline lens and the aqueous and vitreous humours. The lens serves largely for accommodation. We shall understand most clearly how vision is effected by the eye, if we consider how it sees a point object, since an extended object can be analysed into a set of point objects, and what the eye does to one point object it does to the other point objects of which our extended object consists. Let us consider, then, a point object at infinity on the axis of the eye. It sends a set of rays parallel to the axis of the eye on to the cornea and these are converged, chiefly by the refraction at this surface, so as to come to a

TABLE 8

THE MAGNITUDE OF THE OPTICAL SYSTEM OF THE EYE

	Unaccommodated.	Accommodated for 152·5 mm.
Refractive indices :		
Aqueous humour	1·337	1·337
Vitreous humour	1·337	1·337
Crystalline lens	1·437	1·437
Distance from vertex of cornea of :	mm.	mm.
Front surface of crystalline lens . . .	3·6	3·2
Back surface of crystalline lens . . .	7·2	7·2
Fovea	24·0	24·0
Radii of curvature :		
Front surface of cornea	7·83	7·83
Front surface of crystalline lens . . .	10·0	6·0
Back surface of crystalline lens . .	6·0	5·5
Distance from vertex of cornea of :		
First principal focus	13·74	12·13
Second principal focus	22·82	20·95

focus on the retina close to the fovea, which is just to one side of the optic axis (Fig. 97). Therefore a sharp image of the point is thrown on the retina and so the person sees the point distinctly. In the same way he will see an extended object clearly, as each point will send a parallel beam on to the cornea, the direction of each beam being slightly different, and so the position of its point focus on the retina being slightly displaced relative to that of neighbouring beams. In this way the eye forms a sharp image of the object on the retina and the person sees the object distinctly. The image of an erect object will be inverted, but this creates no difficulty as soon as the person has learned by experience to correlate his visual and tactual experience. He then automatically translates an inverted image on the retina to the existence of an erect object external to himself. It is easy to prove that the image is inverted by the following simple experiment. Cast a shadow of a pencil on to a screen with a point source, noting that if the pencil is erect, so is the shadow. Now place a lens just between the pencil and the screen and notice that this does not affect the relations between the object and its shadow. If a small hole is made in a piece of paper, which is then

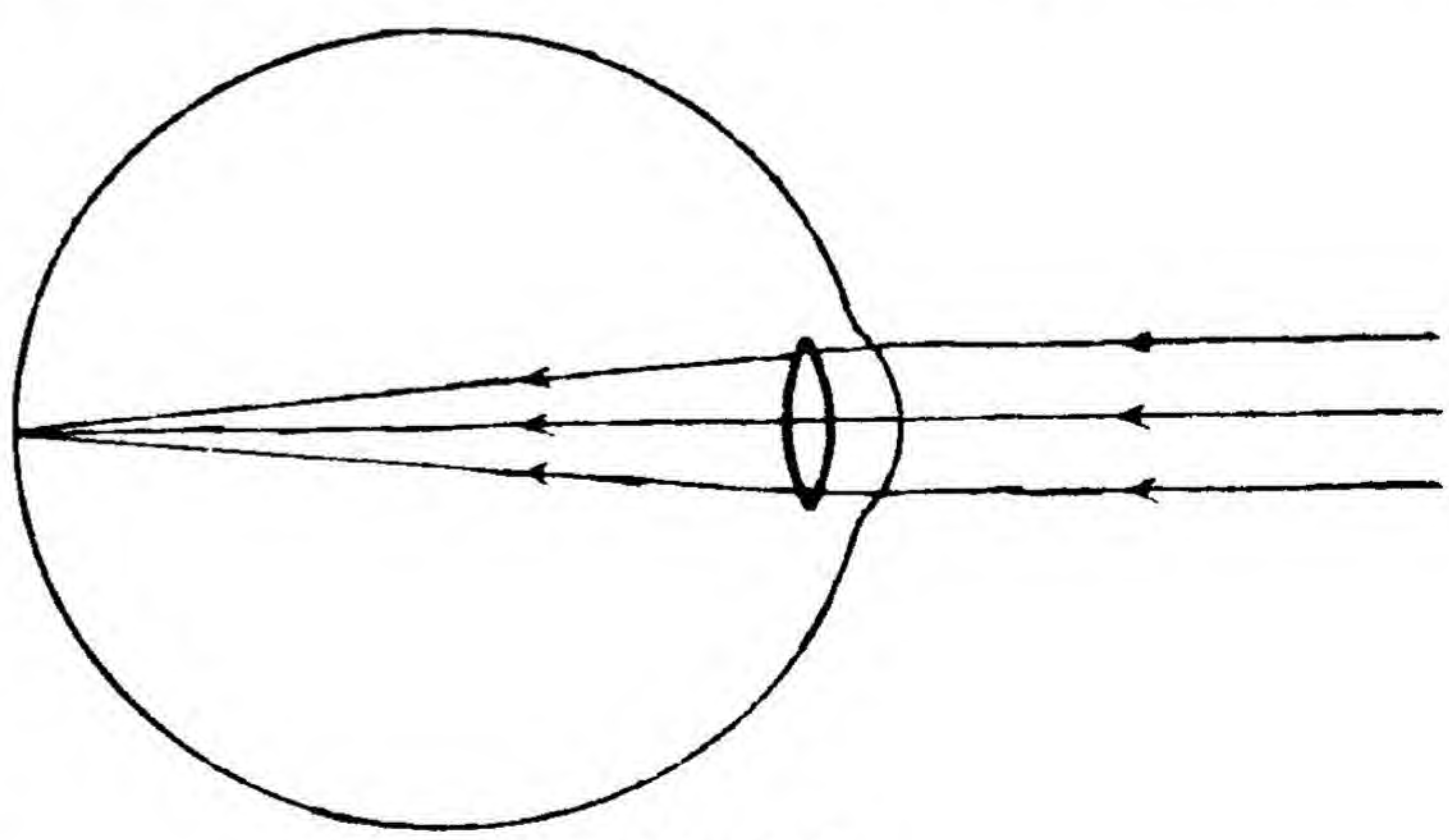


Fig. 97.

held between the eye and a bright light, and a pin with its head uppermost is moved up so as to come between the pin-hole and the eye, a blurred view of the pin with its head downwards is seen. The previous experiment proves that the shadow of the pin cast by the illuminated pin-hole on the retina is erect, but the object seen by the eye is inverted, which shows that the image of an object formed by the eye is inverted.

When the eye sees distant objects, the crystalline lens is at its weakest and the ciliary muscles are relaxed. With this arrangement the second principal focus is at the retina. To see objects close to, as for example when reading a newspaper, the ciliary muscles contract and the curvature of each face of the crystalline lens is increased, thus increasing the power of the lens and so enabling the divergent beam of rays from each point on the newspaper to be brought to its appropriate point focus on the retina (Fig. 98). So the reader can see the print distinctly. This ability to decrease the effective focal length of the eye, or to increase its power, is called **accommodation** and it can be measured in numbers in the following way. Let us imagine that the eye is replaced by a model, in which all the refraction is done by a thin lens placed at the cornea, and that the aqueous and vitreous humours are replaced by air. If we assume that the distance from the lens to the nearest point of the retina is 25 mm. for simplicity, then the focal length of the lens when viewing distant object is 2·5 cm. and its power is 40 dioptres. In order to view an object 10 cm. away, the focal length of the lens must be decreased so that the image of this object is cast on the retina 2·5 cm. away, and so we have from the usual equation

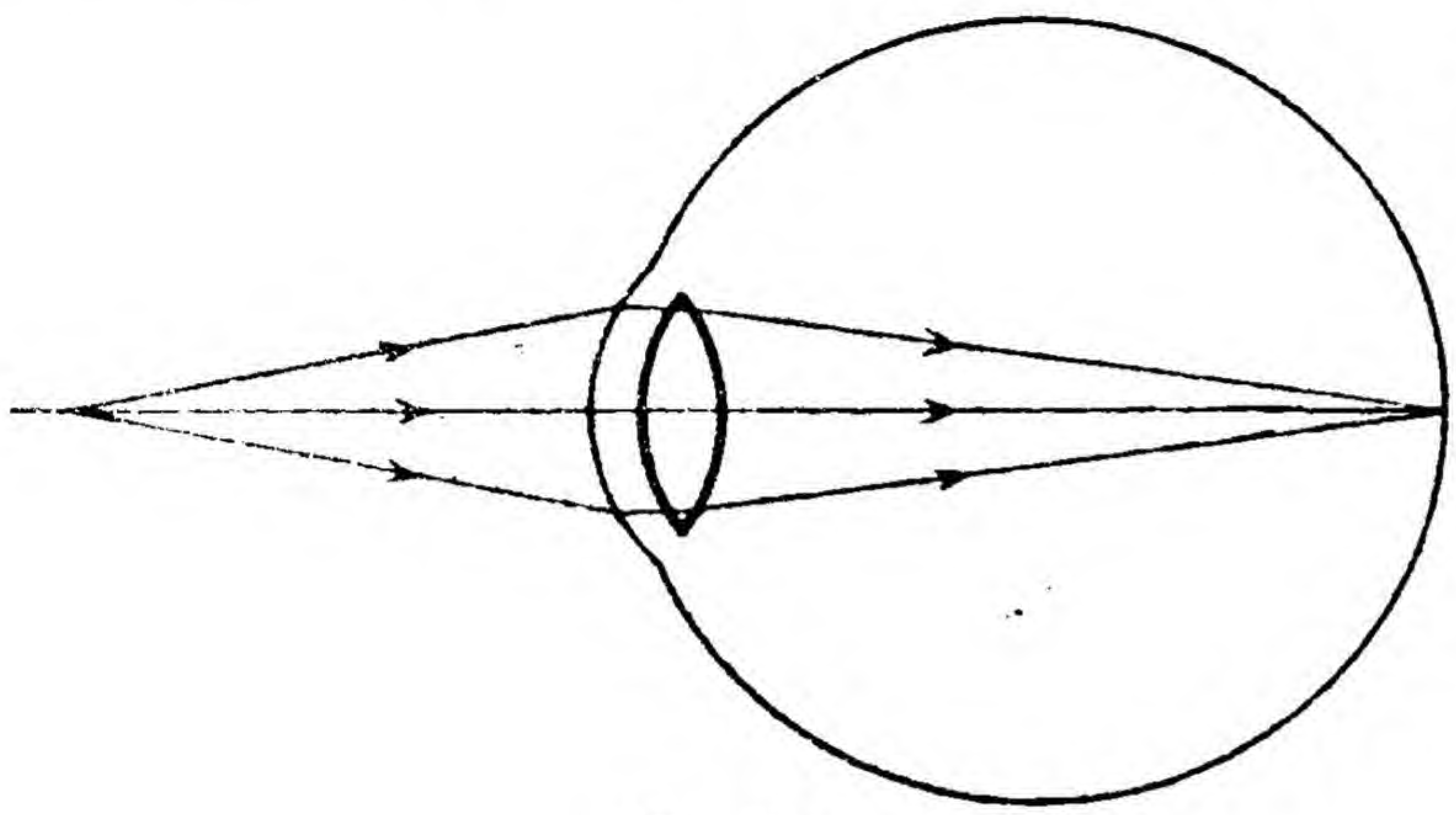


Fig. 98.

$$\frac{1}{+2{\cdot}5}-\frac{1}{-10}=\frac{1}{f}$$

whence $f=2{\cdot}0$ cm.

and so the power of the lens must be increased to 50 dioptres. If this object is at his **near point**, which is the least distance from the eye at which he can see objects distinctly, then his accommodation is 10 dioptres. We see that it is the greatest change in power in the optical system, which can be produced by the ciliary muscles. The way in which the accommodation changes with age is shown in Table 9, which is taken from Hardy and Perrin's " Principles of Optics," and is for a typical eye. This means that it is the mean of a number of observations made on normal eyes. As is to be expected, the accommodation decreases with age and

ultimately may become zero owing to the ciliary muscles losing their power and to the crystalline lens becoming less elastic.

TABLE 9

Age in years.	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
Power of accommodation in Dioptres.	14	12	10	8	7	5·5	4·5	3·5	2·5	1·75	1·00	0·50	0·25	0	0

While considering normal vision we may usefully refer to a number of other interesting points in connection with the eye. The **far point** is the greatest distance at which objects can be seen distinctly and is at infinity for the normal eye. The point at which the optic nerve enters the retina is insensitive to light and is called the **blind spot**. If any part of an image falls on it, that part will not be seen. Its existence can easily be demonstrated by drawing a small circle about an eighth of an inch in diameter and a large dot about three-eighths of an inch in diameter some three inches apart on a piece of paper. The piece of paper is now held some distance away from the eyes, so that the small dot is opposite the left eye and the large one opposite the right eye. If the left eye is closed and the small dot is steadily observed with the right eye, the large dot will disappear if the paper is moved to and from the eyes, until the correct distance away is found. This occurs when the image of the large dot falls on the blind spot and, since the large dot is to the right of the small one, whose image will be formed almost directly behind the centre of the pupil, the blind spot is on the nasal side of this point. We now turn from the place where there is no vision to the place where it is most acute, the fovea. This is the region in the middle of the yellow spot where there are only cones and, when a person wishes to see any object distinctly, he turns the whole eye so that its image falls on the fovea. From the dimensions of the fovea and the other constants of the eye, it can be calculated that an object must only subtend an angle of 0·87° at the eye, if the whole of its image is to fall inside the fovea, and so the whole of objects larger than this cannot be seen distinctly. If the eye is looking at a large object, such as the page of a book, it can only examine carefully a small portion of the page and has to scan the whole page to read it all. It is interesting to see that the eye is subject to the same limitations as the transmitter of a television apparatus, which must scan the subject strip by strip. If the eye can only see a limited field distinctly, how distinctly can it see things in that field ? It must be clearly understood that the indistinctness in the remainder is not due to the image being out of focus, but to the nature of the nerve endings in these parts of the retina. Let us express our question more definitely : what is the least angle which two points may subtend at the eye and yet be seen as separate points ? This angle is called the **resolving power** of the eye ;

the smaller it is, the greater the amount of "grain" or detail the eye will see in a given object. It turns out that the resolving power of the eye is about one minute of arc. This limitation is due to two things. Firstly, if the points subtend a smaller angle than this, their images fall on the same cone, and so the eye registers them as one object. This limitation concerns the grain of the retina itself, as it were. The second one is due to the nature of light itself. When we come to discuss this in detail in the later chapters of the book, we shall see that there are slight departures from the two fundamental laws of rays of light and that the image of a point object formed by a lens free from all aberrations is a small disc, whose radius decreases as the diameter of the aperture of the lens increases. When the two points subtend an angle of less than one minute at the eye, these two discs overlap to such an extent that the eye cannot distinguish the presence of two discs and so it cannot distinguish the two points as separate any longer. If measurements are made of the relation between the acuity of vision and the diameter of the pupil, the results are as shown in Fig. 99. The acuity increases until the pupil diameter is 2 mm., due to the decrease in diameter of the disc formed by the lens. It is then about the same diameter as that of the cones and so further increase in the diameter of the pupil does not produce any increase in acuity, since the size of the cones is the controlling factor now. When the diameter reaches about 4 to 5 mm., the acuity decreases again because spherical and chromatic aberration increase the size of the image. So, at very small diameters of the pupil, the nature of light itself lowers the acuity of vision, at large diameters the geometrical shape of the refracting surfaces keeps it down, and at intermediate diameters the anatomical structure of the retina limits it. This finite value of the resolving power of the eye throws important light on the performance of optical instruments in general and in particular on why the actual lens approximates so closely to the thin lens, even when rays making an angle of two or three degrees with the axis pass through them. In order to subtend an angle of one minute at the eye two points 25 cm. away must be 0·073 mm. apart; in fact, **there is no point in presenting a perfect point focus to the eye, since it cannot distinguish between this and a circle of least confusion of diameter 0.05 mm.** Consequently we see that this imperfection of the eye allows a small but finite departure from paraxial conditions without any deterioration in the definition of the image, and

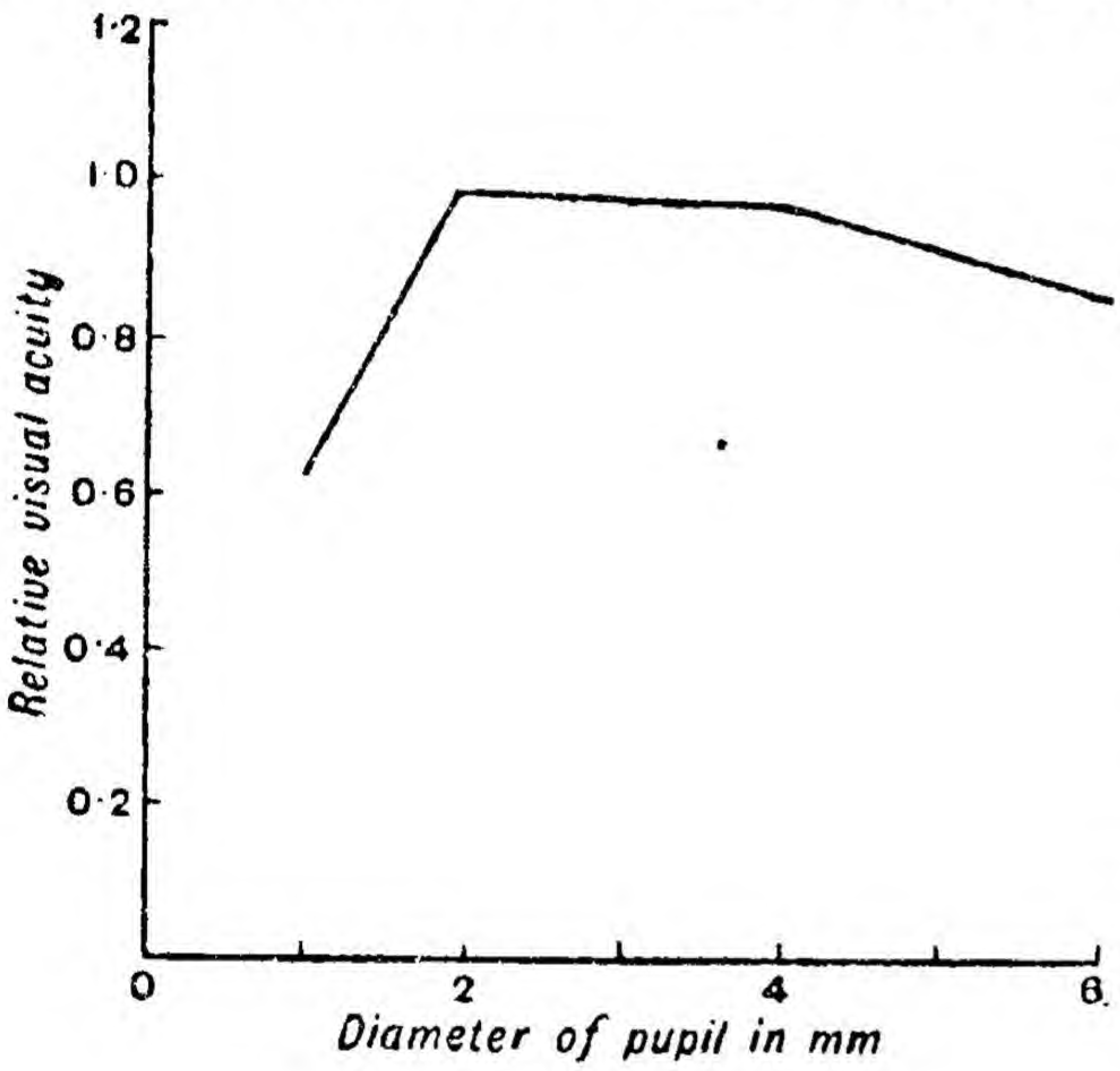


Fig. 99.

this is one of the reasons why an actual lens behaves so nearly like a thin lens and why the concept of the thin lens is so useful in helping us to understand and improve optical instruments (Art. 15).

We have seen that white light consists of a set of homogeneous or pure colours ranging from red through the spectral colours to violet; it is interesting to see how the sensitivity of the eye varies with colour. In other words, if a surface of unit area emits a given amount of energy in unit time, how will the brightness of the surface as estimated by the eye vary as the colour of the radiation emitted is varied from red through the spectrum to violet? This has been investigated experimentally and the results for a typical normal eye are shown in Fig. 100, the curve A being for a surface at normal brightness corresponding to daylight, and curve B being for a surface which is only just visible. If the reader is not familiar with the conception of wave-length as a way of interpreting colour, he may regard it for the moment as a kind of universal refractive index. It will be seen that, under daylight conditions, the eye is most sensitive in the yellow, but that the region of maximum sensitivity shifts to the blue when the surface is much fainter. This is held by some people to account for the fact that moonlight is " cold," since the eye is more sensitive to blue and violet under these conditions.

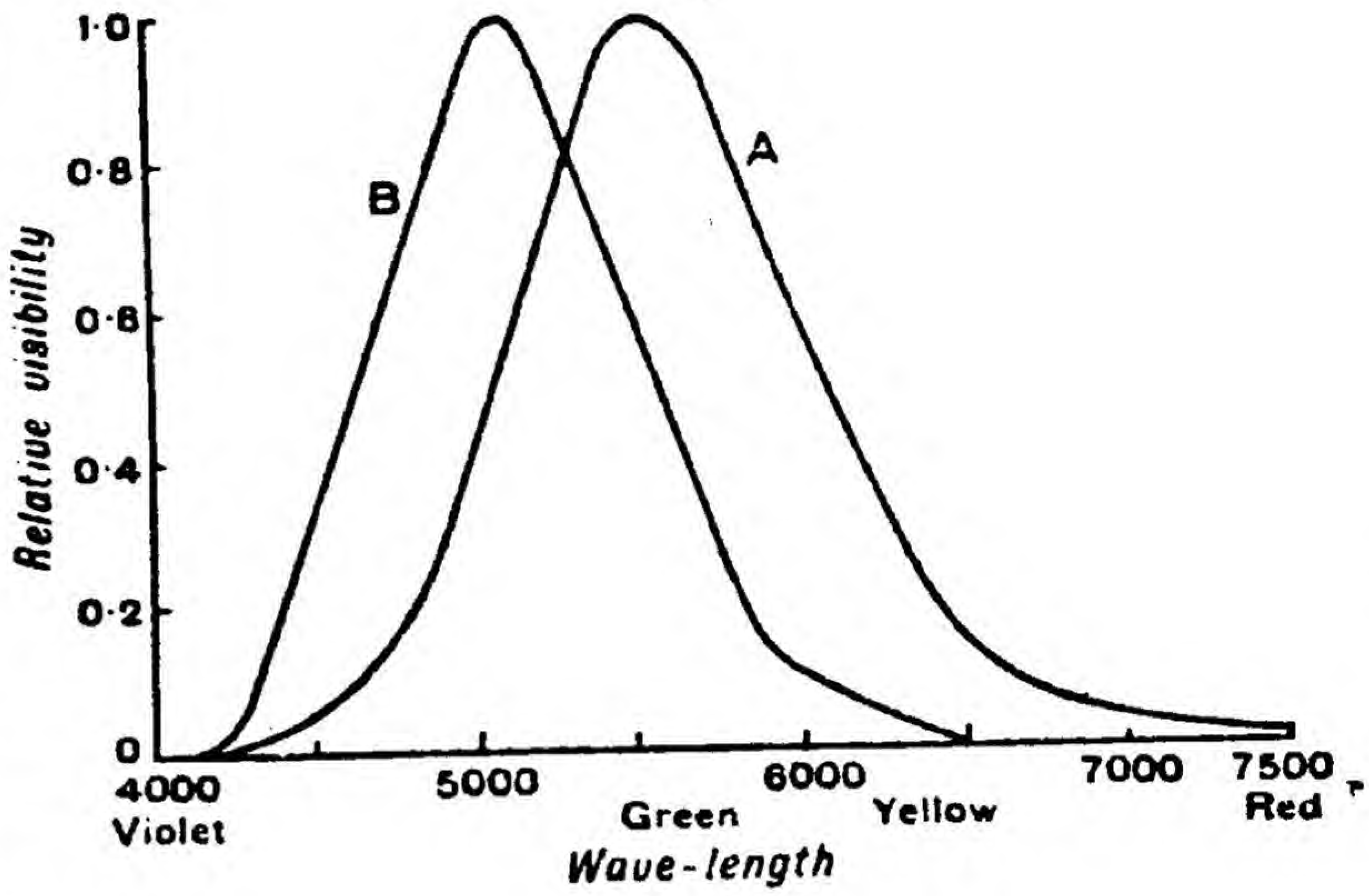


Fig. 100.

56. DEFECTS OF THE EYE

The defects of the eye may be divided into two classes, those due to abnormality in the refracting surfaces of the eye, and those due to diseased condition of some part of the eye. We shall only deal with the first of these two classes, as the second, which includes such diseases as cataract, the formation of an opaque layer over the lens, and detachment of the retina, are matters for the surgeon rather than the physicist. The first defect appears when old age comes on and is called **presbyopia** or **lack of accommodation**. The symptoms are that the subject cannot see distinctly any near objects nor can he read a newspaper, while his vision for distant objects remains good. Also the older he gets, the further away objects must be before he can see them distinctly, or, in technical terms, his near point recedes. This is due to the fact that he loses the power to decrease the focal length of the crystalline lens as he gets older, because

the lens itself loses its elasticity and the ciliary muscles get weaker. We have already given figures illustrating how this power of accommodation decreases with age. The remedy for this defect is to give the subject spectacles consisting of convex lenses for reading, the spectacles performing the converging of the light which should be done by the increase in power of the crystalline lens. If a subject's near point is at 50 cm., the focal length of the lenses needed to restore it to 15 cm. is given by

$$\frac{1}{-50}-\frac{1}{-15}=\frac{1}{f}$$

or

$$f=+21{\cdot}4 \text{ cm.}$$

or about 5 dioptres, since the crystalline lens has lost this amount of its accommodating power. It is obvious that these spectacles bring the far point of the eye from infinity to 21·4 cm., since rays at this distance from the eye emerge from the lens as a parallel beam and will be brought to a focus on the retina by the unaccommodated eye. Consequently these spectacles must only be worn for reading. I have known an engine driver who suffered from presbyopia and who would not go to church because he had to wear glasses to follow the prayers and hymns. He thought that his friends would assume that, as he was wearing glasses for reading, his sight for distant objects was not good and he should not be following his occupation! Needless to say his fears were groundless and, in any case, railway companies test the sight of their drivers every six months.

The second defect is called **myopia** or **short-sightedness**, and the reader will avoid confusion if he realises that the second term refers to what the subject *can* do, not to what he cannot do. The symptoms, then, are that the subject can see distinctly objects a short distance away, but he cannot see clearly objects further away than some fairly definite distance,

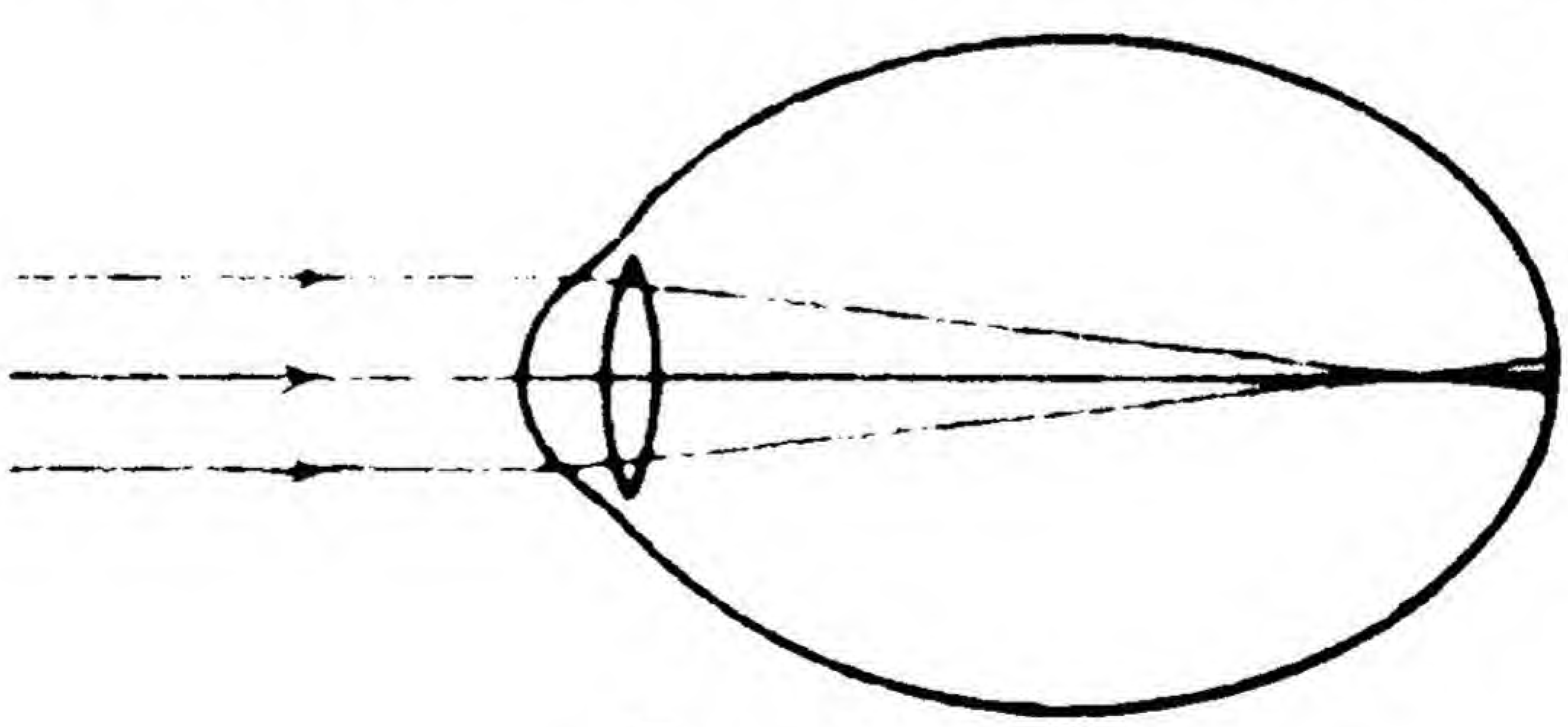


Fig. 101.

which varies for different persons. In technical terms his far point is at a finite distance away instead of at infinity, and his near point is also closer to the eye than is the case with normal-sighted persons. It is quite common for a short-sighted person to have his far point at 50 cm. Instruments are available for measuring directly the radius of curvature of the various refracting surfaces of the eye, and it is found that the radius of curvature of the cornea is practically the same for all eyes. The cause

of myopia is that the eyeball has lost its spherical shape and has become elongated in a direction parallel to the optic axis of the eye and so the retina is abnormally far from the cornea. Consequently rays from a point at infinity on the axis of the eye come to a focus before they reach the retina and produce a spot of finite size on the retina itself (Fig. 101). The subject sees that point blurred and so he cannot focus objects at infinity. His far point is situated at such a distance from the eye, that the rays diverging from it come to a focus on the retina when the crystalline lens is relaxed (Fig. 102). Since the refracting system of the eye is too

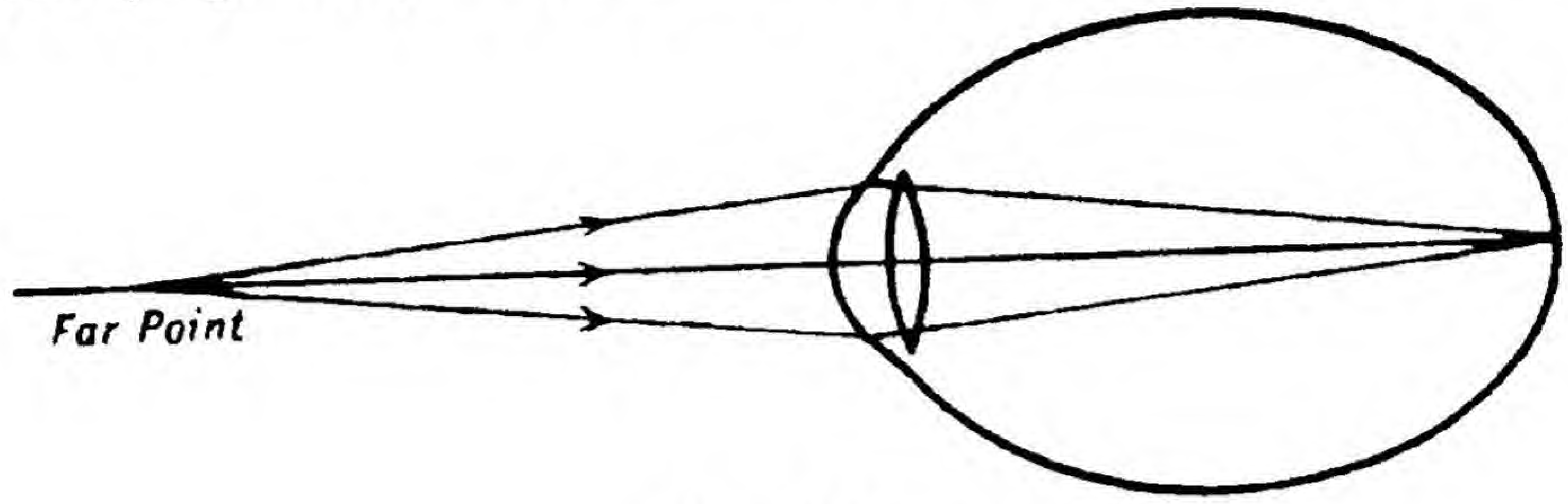


Fig. 102.

powerful relative to the distance from the cornea to the retina, the remedy for this defect is to weaken it by fitting the subject with spectacles consisting of diverging lenses. Lenses of such power must be chosen that the image of a point at infinity on the axis of the eye will be at the far point, and so the typical subject mentioned above will need lenses of power -2 dioptres (Fig. 103). If his near point is at 8 cm., the reader will find it interesting to calculate its position when he is wearing these spectacles. Myopia usually comes on between the ages of ten and sixteen,

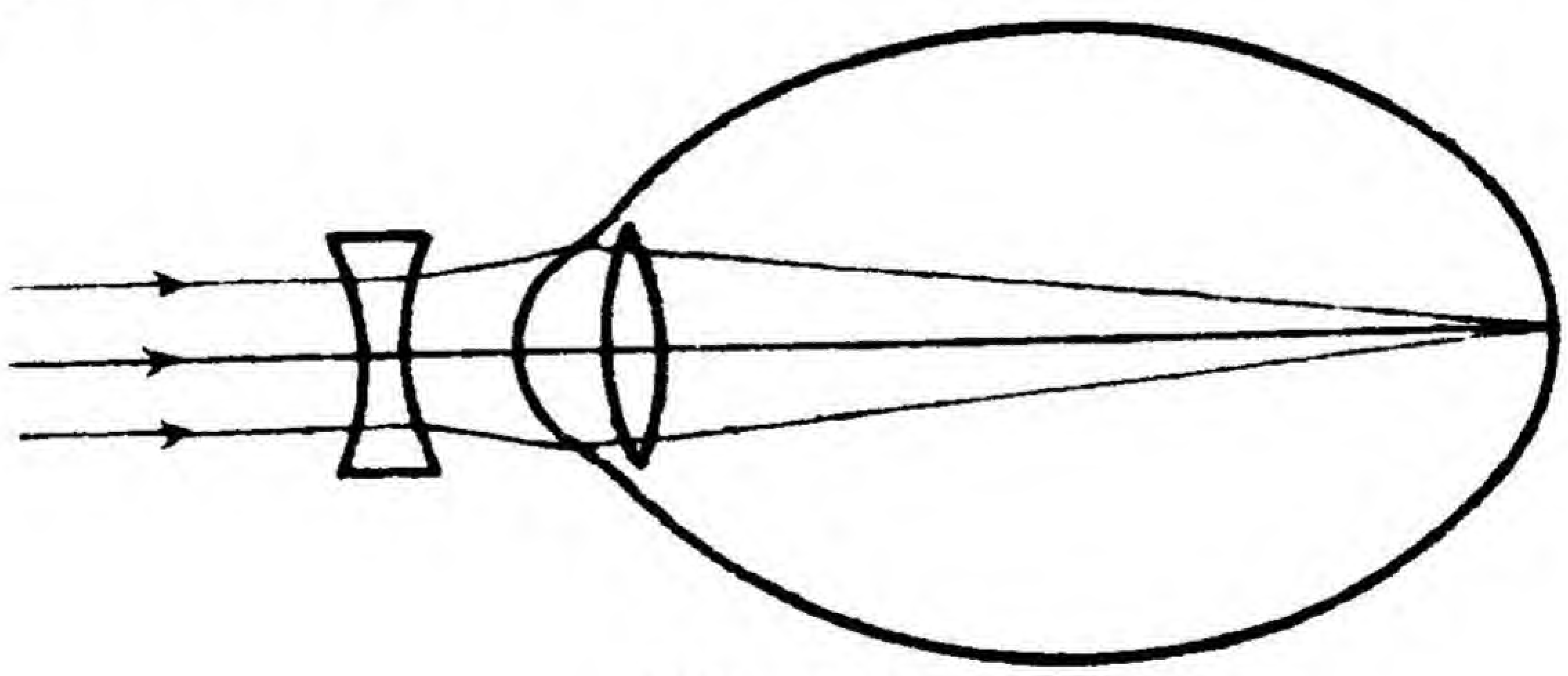


Fig. 103.

when the subject is growing fast and his eyes are developing, and may be caused by too much reading. Apparently the ciliary muscles may strain the shape of the eyeball, if they are used too much. It is being claimed that myopia can be cured by restoring the normal shape of the eyeball by suitable massage. I have known candidates who have failed the medical examination for entrance to the navy and have undergone such treatment in the effort to reach the necessary standard. They have succeeded, but these methods have not yet come into general use and their efficiency is by no means established.

The third defect is **hypermetropia** or **long-sightedness.** The

symptoms are that the subject can see distant objects distinctly, but he cannot see near objects clearly, in other words, his near point is abnormally far from the eye. He also suffers from headaches. The cause of this defect is that the eyeball is elongated in a direction at right angles to the optic axis, and so the retina is abnormally close to the cornea, which, it must be emphasised, has the same curvature as in the normal eye. Consequently rays from a point at infinity on the axis of the eye

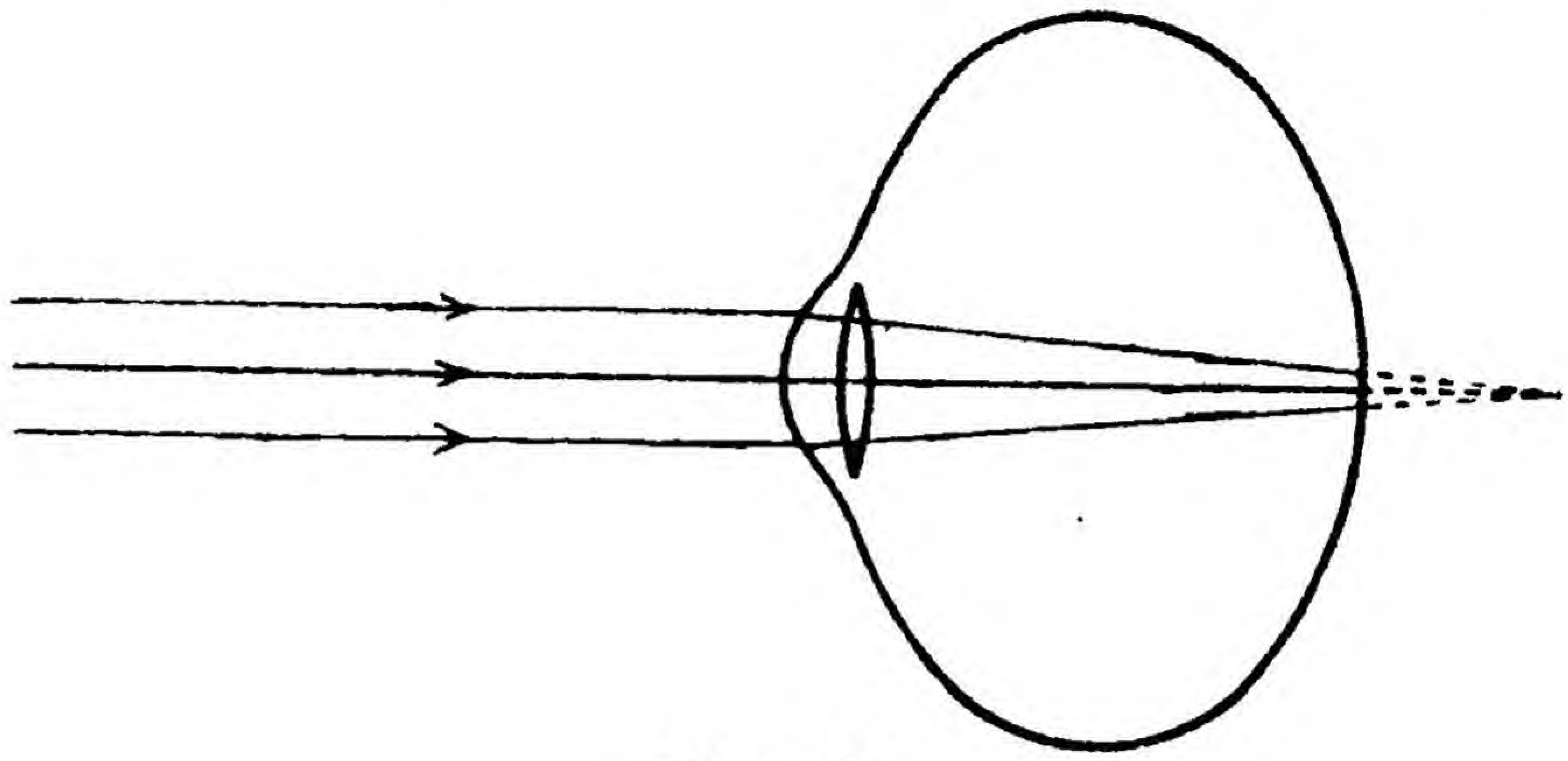


Fig. 104.

come to a focus behind the retina when the crystalline lens is relaxed (Fig. 104), and so the ciliary muscles must always be in use, which is why the subject suffers from headaches. Even when the ciliary muscles have increased the power of the crystalline lens to its maximum, a point on the axis of the eye must still be beyond the normal near point, if the rays from it are to come to a focus on the retina and not behind it. Since the refracting system of the eye is too weak, the remedy for this defect is to

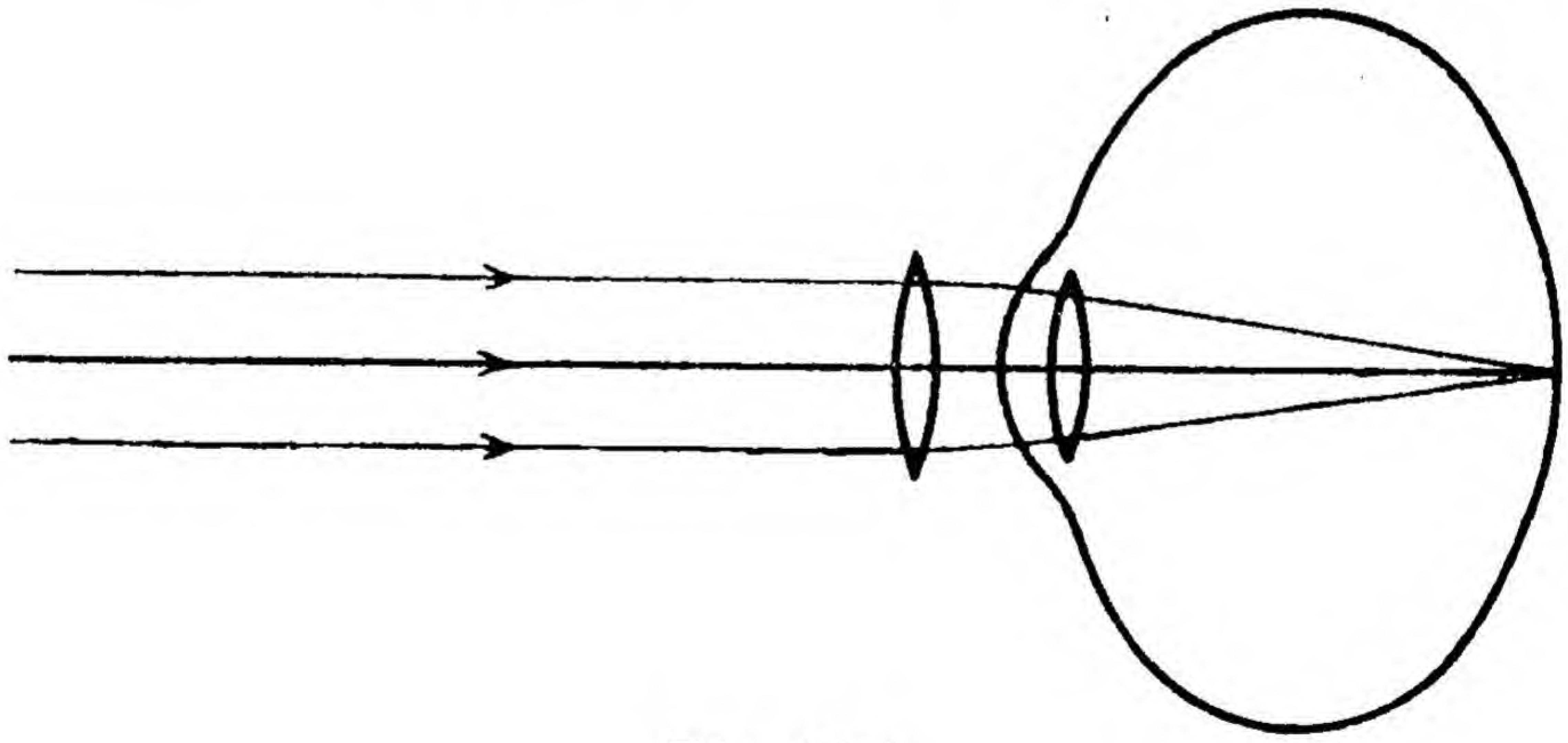


Fig. 105.

increase it by prescribing for the subject spectacles fitted with converging lenses (Fig. 105). If the near point of a long-sighted person is 30 cm. away, then the focal length of the lens needed to enable him to see objects at the more usual near point of 15 cm. is a lens which forms a virtual image at 30 cm. of an object 15 cm. from it. Its focal length is given by

$$\frac{1}{-30}-\frac{1}{-15}=\frac{1}{f}$$

whence $f=30$ cm., and so he needs convex lenses of power $3\frac{1}{3}$ dioptres, which he must wear always so as to avoid having to use his ciliary muscles to see distant objects clearly.

The last and most common defect of the eye is **astigmatism**, and is caused by the curvature of the cornea being different in two mutually perpendicular planes. Since most of the refraction occurs at the cornea when distant objects are being viewed, this means that the focal length of the refracting system of the eye is different for rays in these two planes. Let us imagine, for the sake of argument, that the curvature of the cornea is greatest in a vertical plane and least in a horizontal plane, then the focal length of the eye is least in a vertical plane and greatest in a horizontal plane. This causes the refracted pencil corresponding to a solid incident beam to be astigmatic, and so the best focus of a point object which can be obtained is the circle of least confusion, and a person suffering from astigmatism never sees any object at any distance quite distinctly. It also follows that he can focus the vertical strips of mortar in a wall, for example, and not the horizontal ones, or the horizontal ones and not the vertical ones, but never both sets simultaneously. Nearly everyone suffers from astigmatism to a slight extent, but the planes of greatest and least focal length are not always horizontal and vertical. It is detected by asking the subject to look at a card, on which there are a number of fine black lines arranged like the spokes of a wheel (Fig. 106). One of these will usually look more distinct than any of the others, and the one at right angles to this one will look the least distinct. If this is the case, the subject does possess astigmatism with the axes of greatest and least focal length along the lines which are most and least distinct. The defect is cured by prescribing spectacles fitted with astigmatic lenses, whose curvatures and axes are so chosen that, together with the cornea, they form a refracting system whose focal length is the same in every plane. It is important to realise that such lenses must not only have the correct curvatures, but that they must be fitted in their frames with the planes of least and greatest focal lengths parallel to the directions of greatest and least focal length of the eye. If the lens drops out of its frame, the subject cannot just put it back again, if the frame is circular, but must take it to the optician to have it correctly fitted. It is believed that the famous diarist, Pepys, whose eyesight failed in old age, was merely suffering from astigmatism, which could not be corrected in those days, partly because the nature of the defect was not understood and partly because, even if it had been, the skill necessary to grind cylindrical lenses had not been developed.

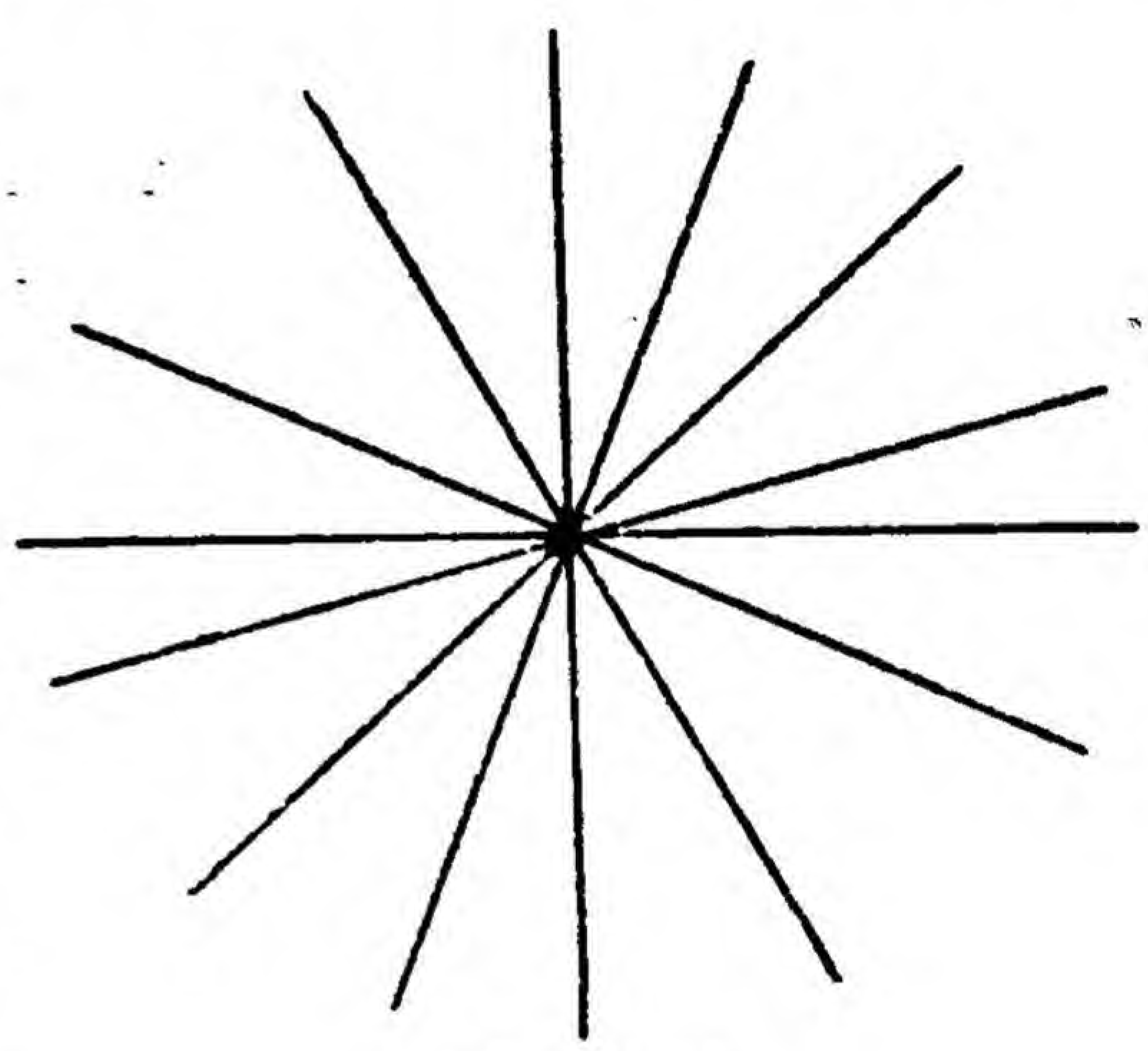


Fig. 106.

57. SIGHT TESTING

We shall now give an outline of the method by which an optician tests a subject's sight, and how he finds the precise power of the lenses needed to correct any defects which he may discover. The reader might be inclined to think, from what he has already read, that the optician will first find the positions of the subject's near and far points. If he finds that the subject's far point is at 50 cm., for example, then he knows that the subject is myopic and needs spectacles fitted with concave lenses of focal length 50 cm. But what does the statement at the far point is at 50 cm. imply? Presumably it means that the subject can read print distinctly at 49 cm. but not at 51 cm. It is safe to say that no such person has ever been seen in an optician's testing-room, and indeed that he does not exist outside the text-book. A simple case of this kind serves as an admirable and simple illustration of the defect and the lines along which it is corrected, but, in practice, it is quite impossible to fix the far point

Z N V T U E

N O H X E Z A U

Fig. 107.

within such close limits as those specified above, and so a quite different method is adopted.

The optician first needs some way of specifying the visual acuity of the subject, and a special set of letters has been devised to do this. They are arranged in horizontal rows on a card, the letters on any one row being of the same size. The standard sizes are such that the letters can be read by a normal-sighted person at distances of 2, 3, 4, 5, 6, 9, 12, 18, 24, 36, and 60 metres respectively, the letters themselves subtending an angle of five minutes at these distances, while the strokes and spaces subtend an angle of one minute (Fig. 107). A card of one of these sets of letters is known as a Snellen chart, and the subject is placed at a distance of 6 metres from the chart, which is well illuminated, and asked to read with the right eye only the row of largest letters first and to continue to the next row and so on, until he reaches the row containing the smallest letters which he can read correctly. If a normal-sighted person could read this row at 12 metres, then the acuity of the subject is $\frac{6}{12}$; that is, his actual distance from the chart divided by the distance at which he should be able to read the smallest letters, which he succeeds in reading

correctly. If the smallest letters which he can read are those which a normal-sighted person could read at 6 metres, his acuity is $\frac{6}{6}$, which is normal; while, if he can read the letters which a normal-sighted person could read at 3 metres, his acuity is $\frac{6}{3}$, which is abnormally high but is found in some people. The left eye is then tested in the same way.

If the acuity of either eye is below normal, the optician then carries out an **objective examination** of that eye in order to discover the cause of defective acuity. A simple instrument which can be used for this purpose is the **ophthalmoscope**, which was invented by Helmholtz (Fig. 108). In its simplest form it consists of a silvered mirror with a hole in the centre, which is used to reflect a beam of parallel light into the subject's eye and so to illuminate his retina. Some light will emerge from the retina out through the cornea and through the small hole in the centre of the silvered mirror and so into the eye of the optician, who is thus able to examine the subject's retina. If the subject's eye is

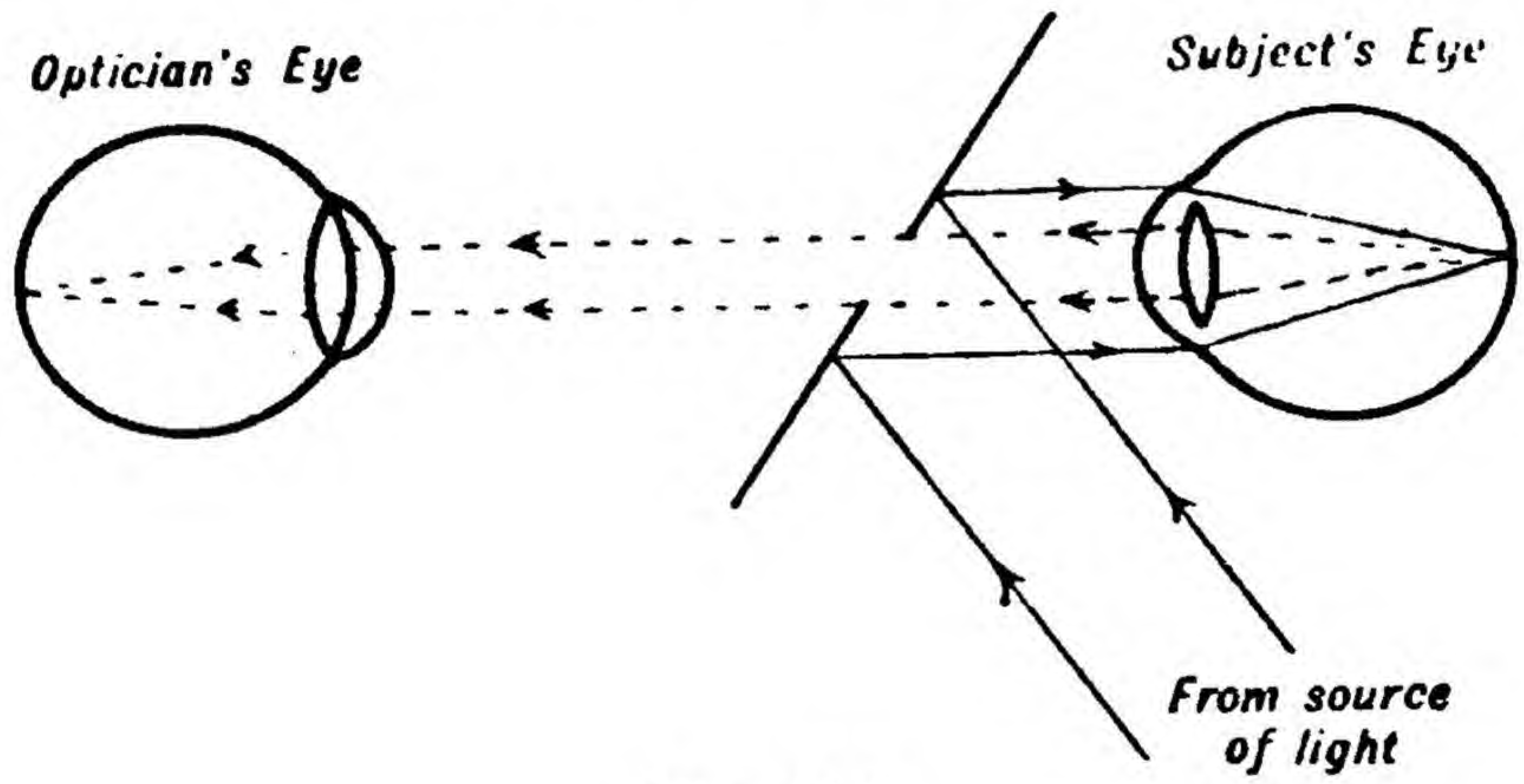


Fig. 108.

normal and is unaccommodated, then the beam of parallel light will come to a point focus on the retina, but if an extended source is used, a finite area of the retina will be illuminated. The beam emerging from any one point of the retina will be rendered parallel by the unaccommodated eye and will enter the optician's eye, which will bring it to a focus on his own retina, if his own eye is unaccommodated. Consequently the optician will be able to see the subject's retina clearly in the ophthalmoscope, if the subject's eye is normal and unaccommodated. Any tendency on the part of the subject to accommodate his eye is prevented, if necessary, by paralysing the ciliary muscles with a drug such as atropine, while a similar tendency on the part of the optician is overcome as the result of continual practice. If, however, the subject is myopic, it is clear that the optician cannot see the subject's retina clearly, since rays from a point on it emerge from the cornea as a convergent beam and come to a focus at the subject's far point, which may be very close to the optician's eye, and, in any case, his unaccommodated eye will certainly not bring such a beam to a focus on his own retina. The subject is then fitted with a trial frame, which fits in front of his eyes just like

a spectacle frame, and into each side of which two or more lenses can be placed. The optician finds by trial and error the weakest concave lens, which will enable him to see the subject's retina clearly. It is clear that this is the lens needed to correct his myopia, since it makes a beam of rays starting from a point on the retina emerge from the spectacles as a parallel beam and so it will also cause a parallel beam striking the spectacles to come to a focus on the retina and so remove the far point of the eye to infinity. Precisely the same thing is done with a hypermetrope, the strongest lens being found which enables the optician to see the subject's retina clearly. Each eye is tested separately in this way. It is right to add that the objective examination is more commonly carried out with a retinoscope, but this method of finding the precise defect of the eye and the power of the lenses needed to correct it is rather involved, and any reader who is interested should consult a book on sight-testing.

This objective examination, so called because the subject is quite passive and is not using the ciliary muscles in any way so as to improve his vision artificially, is now followed by a **subjective examination.** After the effect of the atropine has worn off, the subject is once more put before the Snellen Chart at the standard distance of 6 metres and the trial frame is fixed in front of his eyes. First the right eye is examined in this way ; the lens found by the ophthalmoscope is placed in the trial frame and the subject is asked to read the Snellen Chart. If his acuity is not $\frac{6}{6}$, then other lenses are substituted in the trial frame so as to alter the power of the experimental spectacles, until his acuity is as nearly normal as possible. In the case of a myope, the power of the lens in the trial frame is so chosen as to be a little less than that needed to bring vision up to normal, since myopes tend to hold books too close to their eyes and so to put unnecessary strain on the ciliary muscles. In the case of a hypermetrope, the optician puts a convex lens in the trial frame which is strong enough to " fog " the subject, that is, even with his eye unaccommodated he still cannot see the letters on the chart distinctly. Then weaker lenses are tried until his acuity is normal. When the right eye has been fitted in this way, a similar test is performed on the left eye and the power of the lens required for each eye is simply that of the individual lens in the trial frame when the acuity is normal or as nearly normal as is desirable. In order to carry out this test efficiently, the optician is provided with a set of trial lenses whose powers vary from $+0{\cdot}5$ dioptres to $+10$ dioptres, and $-0{\cdot}5$ dioptres to -10 dioptres in steps of $0{\cdot}25$ dioptres, and this gives him such a variety of lenses to choose from that he can obtain the necessary correction with the use of only one lens for each eye.

Finally the subject is placed before the chart designed to detect astigmatism (Fig. 106) and, after his left eye has been covered up, he is asked to say which line looks the sharpest. Let us assume that the vertical line is the most distinct. Then, as a rule, it means that the curvature of

the cornea of the right eye is greatest in a vertical plane and he is myopic in this plane. For, if this is the case, a beam of rays parallel to the optic axis of his eye will come to a horizontal line focus before the retina and a vertical line focus on the retina. Therefore he will be short-sighted for horizontal lines and normal-sighted for vertical lines. The optician then proceeds to cure the astigmatism by putting concave cylindrical lenses in the trial frame with their axes horizontal, until he has found a single lens which enables the subject to see all the lines in the chart equally clearly. He has then only two lenses in the trial frame, one spherical lens to correct the long or short-sightedness and one cylindrical lens to correct the astigmatism. Having done this for each eye, he proceeds to make up the prescription by writing down the power of the spherical and cylindrical lens in each side of the trial frame. A typical prescription reads as follows, O.D. standing for oculus dexter, or the right eye, and O.S. for oculus sinister, or the left eye :

	SPH.	CYL.	AXIS.
O.D.	+2·0	−1·0	90°
O.S.	+2·0	−1·5	90°

This means that the subject needs for his right eye a spherical converging lens of 2 dioptres and a diverging cylindrical lens of power 1 dioptre with its axis vertical, while for his left eye he needs the same spherical lens together with a cylindrical diverging lens of power 1·5 dioptres with its axis vertical. This prescription is sent to the workshop of the opticians and the way in which it can be made up into a single lens for each eye and the various types of spectacle lens in use will be considered in the next article.

58. TYPES OF SPECTACLE LENS

When the above prescription is received in the workshop, the lens grinder can make up the spectacle lenses in two different ways. In the case of the right eye he can start with a flat piece of glass and grind one face to be a convex sphere of power 2·0 dioptres and the other to be a concave cylinder of power −1·0 dioptres, the lens for the left eye being prepared in a similar way. The radius of curvature of the faces is calculated from the relation

$$\frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

r_2 being put equal to infinity, as all the power is to be produced by the curvature of one face. Tables are drawn from this relation from which the lens grinder can look up the required radius of curvature, given the

required power and the refractive index of the glass he is to use. Lenses made in this way are called flat lenses, and, while they give good vision when the subject looks through the middle of the lens, there is considerable deterioration if he looks through the edges of the lens at objects to one side owing to the aberrations which occur, when pencils pass obliquely through a thin lens. It is found that this difficulty can be decreased by the use of **toric lenses**, in which the face of the lens nearer to the eye is a concave sphere, while the other face is so ground as to correct both the myopia or hypermetropia and the astigmatism. If there is astigmatism, the curvature of this face will be different in two given mutually perpendicular planes and so its surface is **toroidal**, which is why the lens is called a toric lens. The surface of a circular ring, the cross-section of whose material is also circular, is a good example of a toroidal surface. It is found that the best results are obtained by grinding the face near to the eye to have a power of $-6{\cdot}0$ dioptres, and so the other face of the lens needed for the right eye in the above prescription will be ground to have a power of $+8{\cdot}0$ dioptres in a vertical plane and $+7{\cdot}0$ dioptres in a horizontal plane, the corresponding powers for the lens for the left eye being $+8{\cdot}0$ dioptres and $+6{\cdot}5$ dioptres respectively. Special tools have been designed for the grinding of these lenses.

Finally, there are bi-focals, in which each lens of the spectacles consists of two parts, the upper portion, A, being used to correct distant vision, while the lower portion, B, is a lens of greater power to enable the subject to read comfortably (Fig. 109). The aim of bi-focals is to save the subject the trouble of having to change from one pair of glasses to another, when he wishes to read after having been out of doors or to go out of doors after reading, and the necessity for them can be made clear by the following concrete example. A subject has a far point at 200 cm. and his near point at 10 cm. when he is young, and his myopia is so slight that he will probably not notice it and will not wear glasses. As he grows older, his vision for distant objects will get worse probably because the small amount of astigmatism from which almost everyone suffers increases, and his accommodation decreases so that, when he is fifty, his far point is still at 200 cm. while his near point has receded to 40 cm. So he needs convex lenses to correct for his presbyopia, and his distant vision will also need correction by concave lenses. So he is prescribed bi-focals, in which the upper component is a concave lens of focal length 200 cm., which removes his far point to infinity, and the lower component is a convex lens of focal length of 40 cm., which will bring his near point from 40 to 20 cm., as the reader should verify. It is interesting to calculate his near point when

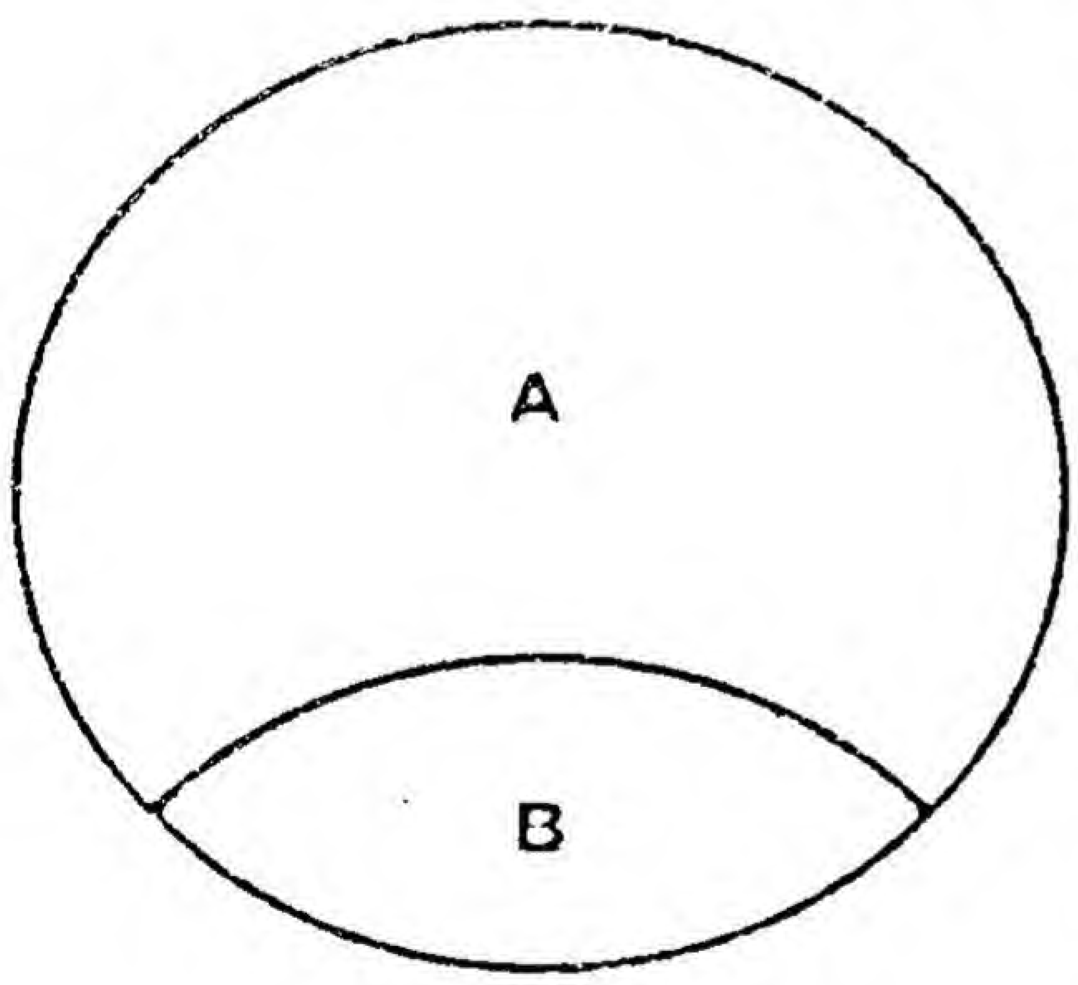


Fig. 109.

using the upper component and his far point when using the lower component. The near point is at a position u given by

$$\frac{1}{-40}-\frac{1}{u}=\frac{1}{-200}$$

from which $u=-50$ cm. The far point when reading is given by

$$\frac{1}{-200}-\frac{1}{u}=\frac{1}{+40}$$

from which $u=-33{\cdot}3$ cm. Therefore the subject cannot see objects distinctly between 33·3 and 50 cm. away when wearing his bi-focals, so, even with these more elaborate spectacles, he still suffers from a kind of blind region, although its extent is much less than that under which he laboured without any spectacles or with only single-lens spectacles.

But it would be quite wrong to conclude this account of the defects of vision and the way to remedy them on a note of pessimism. Lenses and mirrors have been known from the earliest times, and the first spectacles probably arose from the use of the reading-glass, which was merely a convex lens used as a magnifying glass by old people suffering from presbyopia. It was the very existence of mirrors and lenses from earliest times, and the discovery that they produce magnified images and can be used as an aid to vision, which stimulated interest in the theory of optics and led to the discovery of the various theorems of geometrical optics which have been discussed in the previous chapters. Here the knowledge of striking, interesting, and useful facts stimulated pure scientific discovery, the fruits of which have proved useful in industry, in that they have led to a much wider understanding of vision and the correction of its defects. This is surely one of the greatest benefits which science has conferred on humanity, when we reflect on the misery and limitations imposed on persons suffering from short-sight, long-sight, or astigmatism with no hope of any relief. But it is well to bear in mind that the present state of our knowledge in this field is indebted both to pure science and industry. It was the latter which made use of the existing lenses as spectacles ; it was the scientist who devised the explanation of how the spectacles did their task and so explained to the man of industry how to make spectacles to cure long-sight and short-sight as well as presbyopia. He finally mastered the more subtle defect of astigmatism, but here the information derived by the scientist would have been of no avail if the industrial worker had not learned how to grind the new cylindrical and toroidal surfaces required for the correction of astigmatism. It is a mistake to regard the advance in the control and understanding of nature as due entirely to pure science or equally to industry ; the rate of progress is most rapid, when there is two-way traffic between these two spheres of activity, and each becomes sterile if divorced from the other. The force of this argument will become stronger as our investigation proceeds.

59. COLOUR, COLOUR VISION, AND COLOUR PHOTOGRAPHY

While considering the eye as an optical instrument, it is natural to inquire how it distinguishes different colours from one another, which leads to the subject of colour vision and colour blindness, yet another defect of the eye. The word " colour " is used in three different senses, which must be carefully distinguished. It may refer to an object, the red wool, to the objective stimulus proceeding from it, the red light, and to the sensation produced by the stimulus, the sensation of red experienced by a person looking at the red wool in daylight. The term " red light," or " coloured light," will be used to refer to the objective stimulus and the term " red sensation," or " coloured sensation," will be employed to denote the physiological sensation experienced when the stimulus strikes the retina.

Newton established the fact that, while a given coloured light of a given refractive index in a given medium always produces the same colour sensation, that colour sensation is not necessarily due to that same coloured light, but may be due to a combination of various coloured lights. For example, the sodium D light invariably produces the yellow sensation, but the same sensation is not invariably due to sodium D light ; it can be produced by the addition of red and green light in suitable proportions. A large amount of work has been done in recent years on colour mixing along the following lines. One half of the field of view is illuminated with monochromatic light of given wave-length and at a fixed energy, while the other half is illuminated with three given coloured lights of wave-length λ_1, λ_2, and λ_3, whose energies can be varied by a measurable degree. (The reader is reminded that wave-length is a kind of universal refractive index, an absolute way of specifying a given coloured light.) The monochromatic light is first adjusted to have a wave-length λ_1, keeping the number of ergs emitted per second always the same. Then the energies of the wave-lengths λ_1, λ_2, and λ_3 are adjusted so that the colour sensations in the two halves of the field are the same, that is, so that a perfect colour match is obtained. The energies for λ_2 and λ_3 are obviously 0 and that for λ_1 is fixed for convenience at 100. The same procedure is then adopted for λ_2 and λ_3. After this the monochromatic light is varied in wave-length from the violet to the red end of the spectrum keeping the energy output constant, and the energies of each of the wave-lengths λ_1, λ_2, and λ_3 needed to obtain a perfect colour match for each wave-length are determined. Two significant results are obtained by experiments of this sort. Firstly, it is found that the colour sensation of any given monochromatic light can be perfectly matched by the addition of suitable energies of the three coloured lights of wave-length λ_1, λ_2, and λ_3, and a typical set of results is shown in Fig. 110. Secondly, there are not just three coloured lights with which matching can be accomplished,

but many sets of three. These results are of the greatest use in industry, for it is evident that they can be made the basis of an exact specification of any given colour sensation. It is necessary to settle by international agreement what three wave-lengths shall be used for the matching purposes, and once this has been done any given colour sensation can be specified by the energies of these three wave-lengths needed for perfect matching. This will ensure that any particular colour sensation can be scientifically specified and produced at will, a result which could never be guaranteed before the above facts had been established. A detailed

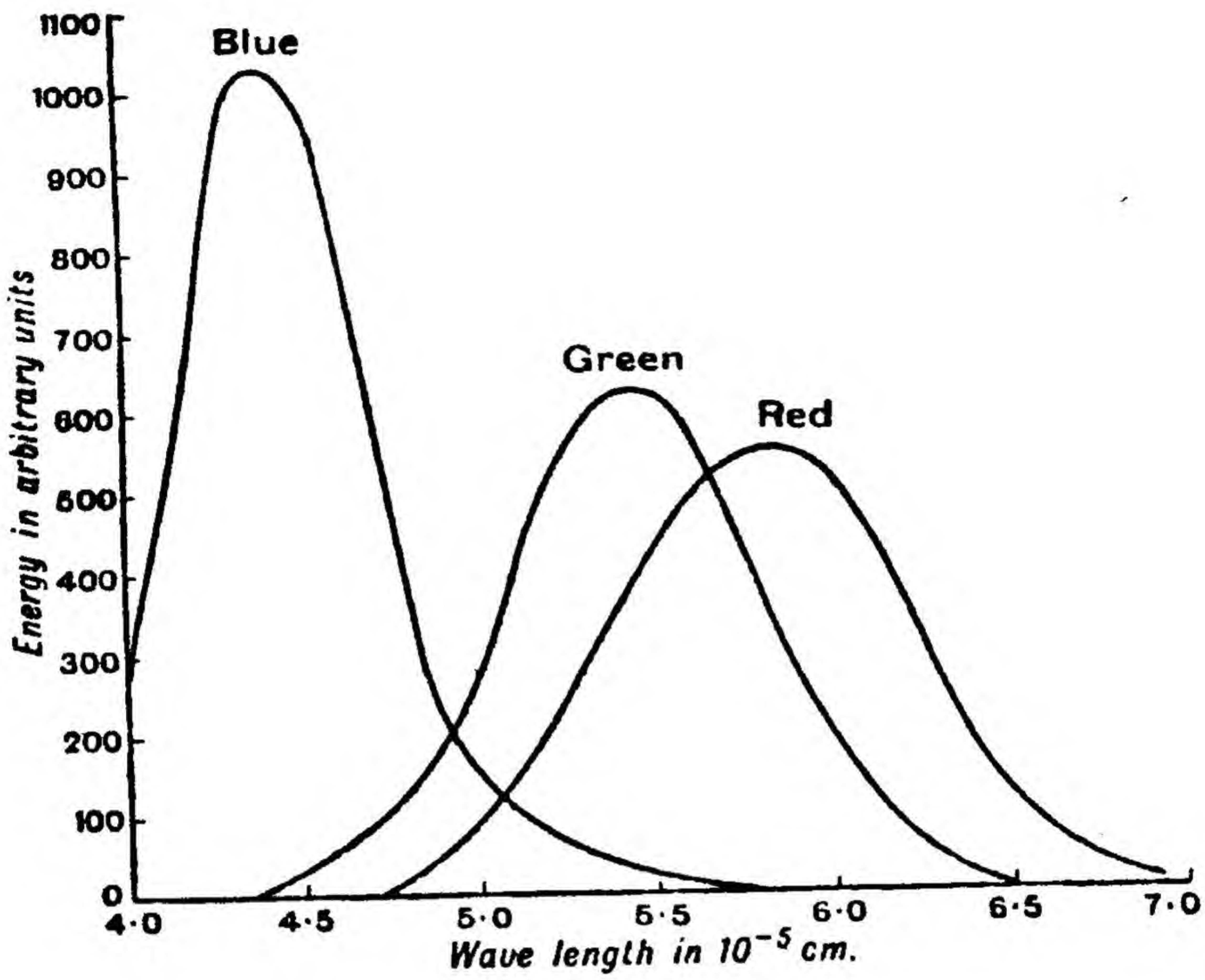


Fig. 110.

treatment of this important advance is out of place in this book, and for further particulars the reader should consult an advanced text-book such as Hardy and Perrin's "Principles of Optics."

What light do these facts throw on the mechanism of colour vision? At the outset, the contrast between the ear and the eye is striking. It is not possible to produce the sensation corresponding to a particular note, such as middle C, by producing simultaneously three other vibrations of suitable frequency. In fact, if three vibrations of different frequency are produced simultaneously and the waves set up enter the ear, it records simultaneously the three sensations which would be perceived, if the three vibrations were produced in turn. If two vibrations of very nearly the same frequency round and about 256 c.p.s. are produced, the ear will record two separate sensations until the frequency difference is less than 2 c.p.s. This ability of the ear to distinguish between notes of nearly equal frequency over the whole audible range is explained by the existence in the ear of a large number of stretched

fibres or strings, whose natural frequencies cover the range of audible frequencies and which respond to the incoming waves. Simplifying matters for the sake of brevity, if the frequencies of the two incoming waves differ by more than 2 c.p.s., two separate fibres are set into vibration and so two sensations, or two notes, are recorded by the ear. Similarly the inability of the eye to register two sensations, when two coloured lights are received, indicates that there are far fewer receiving systems in the eye than in the ear. The fact that the colour sensation corresponding to any given monochromatic light can be produced by the addition of suitable energies of three given monochromatic lights suggests that there are just three sets of mechanisms in the eye, which, when set into operation or excited by incoming light, produce the sensations corresponding to these three monochromatic lights. The extensive evidence which has been collected on colour matching and colour blindness has made it possible to select the three monochromatic lights, whose sensations are experienced when the three sets of mechanisms are excited, out of the many sets of three monochromatic lights, which will produce perfect colour matching. One mechanism produces a red sensation, one a green sensation, and the third a blue sensation. It is clear that these mechanisms must respond to incoming monochromatic light in accordance with the curves shown in Fig. 110, which have been drawn up to represent the results of colour matching by the three monochromatic lights corresponding to the three mechanisms in the eye. As far as the experimental results are concerned, the ordinates represent the energy of each monochromatic light in the mixture, but the units are different for each coloured light and have been so chosen that equal amounts of each light are required to produce the sensation of white light. In practice, this means that the energy of the blue light is greatly exaggerated in this diagram, because it does produce a profound effect on the final colour sensation, although it does not affect the total brightness very much. When these curves are applied to the mechanisms in the eye, the meaning of the ordinates may be understood by considering what happens, when green light of wave-length $5{\cdot}50 \times 10^{-5}$ cm. enters the eye. According to the experimental results, the same sensation as this light is produced by the addition of 20 units of blue light, 610 units of green light, and 430 units of red light. According to the theory of colour vision, the green light produces the sensation it does because it causes the blue, green, and red mechanisms to be excited as if they had been stimulated by the above amounts of blue, green, and red light respectively. In other words, on the theory of colour vision the ordinates measure the extent to which the mechanism responds to a given wave-length, and the curves themselves are really the frequency-response curves of the mechanisms, and they suggest that they are vibrating systems with a natural frequency at the point where the response is a maximum. The facts of colour blindness support this theory of colour vision. They are

briefly as follows: the commonest kind is red colour blindness, in which the sufferer is unable to distinguish red objects from green ones, the spectrum being abnormally short at the red end. It is believed to be due to the sufferer having only two sets of mechanisms instead of three; it would be expected that he would match any colour sensation with a suitable addition of only two monochromatic lights, which is verified experimentally. There are also cases in which the sufferer is blind to blue colours, and finally cases in which the sufferer has no distinction as to colour at all, the whole spectrum looking white. This very brief summary of the theory of colour vision may be concluded with two observations: firstly, there is so far no anatomical evidence for the existence of three sets of mechanisms in the eye and so this theory is by no means universally accepted. A decision must be deferred until further evidence has been accumulated. Secondly, the two principles concerning colour matching are experimental facts and so quite independent of any theory of colour vision. And it is pertinent to emphasise how the ability to measure things in numbers has had yet another triumph, in that it has led to the possibility of an exact scientific specification of colour sensations.

The aim of colour photography is either to project an image of an object in its natural colours on to a screen by sending light through some sort of transparency, such as a lantern slide, or to print an image of the object in its natural colours on paper. An alternative to the first type of process is to produce a transparency, which will show an image of the object in its natural colours when seen against a bright background such as a clear sky. There are two distinct ways in which these things can be achieved; one is an **additive process** and in the other the colours are obtained from white light by a **subtractive process.** We shall describe an additive process for the production of transparencies and a subtractive process for three-colour printing. Maxwell laid down the fundamental principles of the additive process for making transparencies in a Royal Institution lecture in 1861. Let us imagine that we wish to project an image of a blue hyacinth with green leaves in a red pot standing against a white background on to a screen in natural colours. We take three separate negatives of the object, one through a red filter, one through a green filter, and the third through a blue-violet filter. We develop the negatives in the usual way, print positives on glass to act as lantern slides, and project each positive through the same coloured filter as was used in making the negative from which it was derived. If the three images are made to overlap exactly, the image on the screen has the same colours as the object.

To see that this is so, let us consider the lantern slide derived from the negative taken through the red filter. This filter will transmit light from the red pot and the white background only, so that an image of these two objects will be formed on the photographic plate and will reduce the silver

bromide over the corresponding regions. The process of development will produce an image in black silver of the red pot and white background, the remaining silver bromide being removed by the fixing of the plate. Light is now sent through this negative and a lens produces an image of it on another photographic plate. The silver bromide is unaffected where the image in black silver was formed in the first case. When the plate has been developed and fixed, there is a transparent image of the red pot and white background, the rest of the plate being black. This is the first lantern slide. Similarly the lantern slide derived from the negative taken through the green filter will consist of a transparent image of the green leaves and the white background, and the lantern slide derived from the negative taken through the blue-violet filter will be a transparent image of the blue flower and the white background. The three projectors with their appropriate filters are now set up and produce a red patch, a green patch, and a blue-violet patch respectively on the screen. These three patches are made to overlap and their intensities are adjusted, so that a single patch of white light is produced. The lantern slide derived from the negative taken through the red filter is now inserted in the projector provided with the red filter, and it screens red light from the whole of the white patch on the screen except the image of the red pot and the white background. The insertion of this slide changes the white patch into a blue-green patch containing a white image of the red pot and the white background, since blue-green is the colour complementary to red. The lantern slide derived from the negative taken through the green filter is now inserted in the projector fitted with the green filter, and screens green light from the whole of the patch except the image of the green leaves and the white background. The blue-green patch is therefore changed into a blue-violet patch containing a magenta image of the red pot, a blue-green image of the green leaves, and a white image of the background, since blue-violet is the result of subtracting red and green from white light and magenta is the colour complementary to green. The third lantern slide is inserted in the remaining projector and screens blue-violet from the whole of the patch except the image of the blue flower and the white background. The blue-violet patch is changed into a black patch containing a red image of the red pot, a green image of the green leaves, a blue-violet image of the blue flower, and a white image of the white background. The reason for this is that red is produced by subtracting blue-violet from magenta, green by subtracting blue-violet from blue-green, and the background appears white because its image has not been screened from any of the lanterns. So an image of the object in natural colours is cast on the screen and the white background appears white in the image, because the three colours forming it—red, green, and blue-violet—are present in such proportions as to produce white. This is the reason why this method is called the additive process. Any colour in an object other than the three primary colours is produced

by their addition in a similar way to that in which white has been produced in the above case.

This process was difficult and cumbersome in Maxwell's time, because the filters were tanks of liquid. Also it is necessary to take three separate negatives of the object and, if it moves in between any of the exposures, the result is spoiled. The results have been greatly improved in recent years by the production of filters consisting of thin sheets of gelatine impregnated with suitable absorbents and by the manufacture of a mosaic incorporating the three filters, which obviates the necessity of taking three separate exposures. A piece of Dufay colour film consists of a strip of celluloid, on which is deposited a "reseau" of red lines interspersed with rows of alternate green and blue-violet spots. The emulsion containing the sensitive silver bromide is placed on the top of the reseau and is covered with black paper. Let us imagine that we are photographing the same object as before. The camera lens casts an image of the object on the film, the light passing in succession through the celluloid, reseau, and emulsion. Let us think of the image of the red pot. Since only red light is falling on the film over this region, the emulsion under the red lines only will be affected, since the red light is absorbed by both the green and blue-violet dots. The emulsion under the green dots only is affected over the image of the green leaves, and that under the blue-violet dots over the image of the blue flower. The black paper is now stripped off the emulsion and the film is developed, but not fixed. The silver under the red lines, where the image of the red pot is situated, is now bleached away, and the same thing is done for the other regions of the film where there is silver, leaving these regions transparent. The rest of the film is now developed in the ordinary way, the silver bromide becoming black. Therefore the film is transparent under the red lines in the region of the image of the red pot, being black under the green and blue-violet dots; similarly it is transparent under the green dots where the image of the green leaves has been formed, being black under the red lines and blue-violet dots; lastly it is transparent under the blue-violet dots where the image of the blue flower is situated and black under the red lines and green dots. Therefore, if white light is sent through the film, only red light can pass through it in the neighbourhood of the image of the red pot and, if an image of the film is being cast on a screen, an image of the pot in red is formed. No grain or detail is seen in this image, as the red lines in the reseau are too close together to be distinguished as separate. In the same way, only green light can pass through the part of the film where the image of the green leaves was formed and so they are projected in their proper colour, or will look green, if the film is viewed against a bright background, such as the sky or a white light. The same thing is true of the blue flower. Finally the white background looks white, since the corresponding region of the film is affected under the red lines, green dots, and blue-violet dots and is transparent under all those dots after the developing,

bleaching, and second developing processes have been carried out. Hence red, green, and blue-violet light emerges from the parts of the film, where the image of the white background was formed, and, as these colours add to white, the background appears white on the screen by addition of the three primary colours.

We shall now illustrate the subtractive process by describing the production of a coloured picture of the above object by colour printing. Three negatives of the object are made, one through a red filter, one through a green filter, and the third through a blue-violet filter. Positives are then printed on white paper of each of these negatives, but, instead of printing them in white and black, the negative taken through a red filter is printed in white and the colour complementary to red, minus red or blue-green, the one taken through a green filter is printed in white and the colour complementary to green, minus green or magenta, and the one taken through a blue-violet filter is printed in white and the colour complementary to blue-violet, minus blue-violet or yellow. These positives are laid down on white paper on top of one another, the one printed in white and yellow being next to the paper, the magenta next, and the blue-green on top. Now consider the red pot; it sent no light on to the negatives taken with the green and blue-violet filters, and so the prints from these negatives will produce a magenta and yellow image of the pot, while the print from the third negative lays nothing on the paper at the image of the red pot. When white light falls on to the white paper covered with the three prints, the light falling on the region of the image of the pot passes through the magenta print, which robs it of green, and then through the yellow print, which robs it of blue and violet, so that the only colour left is red. The process is intensified on the way back from the white paper again and only red light finally emerges; therefore the red pot appears in its natural colour, red. The reader should work out for himself how the leaves appear green, the flower blue, and the background white. This is called the subtractive process because the red colour, for example, is obtained by subtracting all the colours of the spectrum but red from white light. It can also be used in producing transparencies by printing positives from the three negatives on to a transparent film, or glass plate, instead of on white paper.

60. CAMERA LENSES

The camera consists essentially of a lens mounted in the front face of a light-tight box and a flat plate or film, sensitive to light, which fits against the back face. The distance of the lens from the plate is adjusted so that it casts a real, inverted, and diminished image of the object to be photographed on the plate, which makes a permanent record of the image, from which the final photograph is produced. It will be noticed that the plate plays the same part in the camera as the retina in the eye, but that its record is a permanent one, whereas the retina ceases to register

anything in about 0·1 seconds after the light ceases to fall on it. The camera lens plays the same part as the cornea and crystalline lens of the eye, but the ability to focus objects in different positions is obtained by moving the lens in the case of the camera. But the demands made on the camera lens are much more severe than those made on the eye ; the eye only covers a field of view of 0·87° for distinct vision, whereas quite cheap cameras cover a field of 50° ; the diameter of the aperture of the refracting system of the eye expressed as a fraction of the focal length is f/6 under average conditions of illumination, while camera lenses for Press photography have apertures up to f/2 ; and the definition of the image produced by the camera lens must be superior to that produced by the eye, as it frequently has to be capable of considerable enlargement. The design of such lenses is a matter of considerable skill and the various aberrations of a thin lens are reduced to tolerable amounts by the methods outlined in Chapter 6. It is important to realise that, even if the "ideal lens" could be made, which would produce a perfectly sharp flat image of a large plane object, it would be no use to the photographer. He has to take photographs of solid objects, not flat ones, and such a lens would give a sharp image of the portions of the object lying in one plane, while the rest of the object would be more or less blurred. If the object was a person's face, for example, and the camera was focussed on the ears, then they would stand out sharply while the cheeks, eyes, and nose would be progressively more blurred. It is quite certain that no one would buy a photograph of this kind ! But actual lenses do not produce a point focus of a point object, but a circle of least confusion, and in any case the eye cannot distinguish between a point and a circle of diameter 0·1 mm. Furthermore the least size of the circle of least confusion, which the lens need produce to obtain the best definition, is increased in photography by the finite size of the grains of the sensitive material with which the plate or film is coated and it is found by experience that there is no point in reducing the diameter of this circle below 0·25 mm. This fact ensures that the lens can throw on to the plate a sufficiently well focussed image of a solid object, since the circle of least confusion produced of points in the plane on which the camera is focussed is less than the above minimum, so keeping the size of the circles produced in the other planes below the minimum. The difference in distance from the lens of the nearest and furthest planes which are well focussed is called the **depth of focus** of the camera and the factors which control it will be discussed later.

Cameras can be divided into four classes according to the performance of which they are capable. The first class is the cheap fixed focus camera with an aperture of f/11, so that snapshots can only be taken in bright sunlight, covering a field of 50°, and giving fair definition towards the edges of the field. The second class is also used by amateurs, and has a movable lens with an aperture up to f/4·5, so that snapshots can be taken

in dull weather; it covers a field of 50° and gives good definition right to the edges of the field, so that enlargements are possible. The third class consists of cameras used by the Press, in which the aperture goes up to f/2 to enable high speed photographs to be taken, and the definition is very good over a wide field, as enlargements are often needed. The fourth class consists of cameras needed for special purposes, such as those used in aerial surveying, for taking portraits, wide-angle cameras used for covering fields up to 120°, and telescopic cameras used for taking photographs of athletic events from the grandstands. We shall now consider very briefly how lenses have been designed to meet the requirements of these various classes.

The fixed-focus camera consists of a single meniscus lens with a stop of fixed diameter giving an aperture of f/11 mounted in front of it (Fig. 111). Chromatic aberration is not corrected for; the position of the

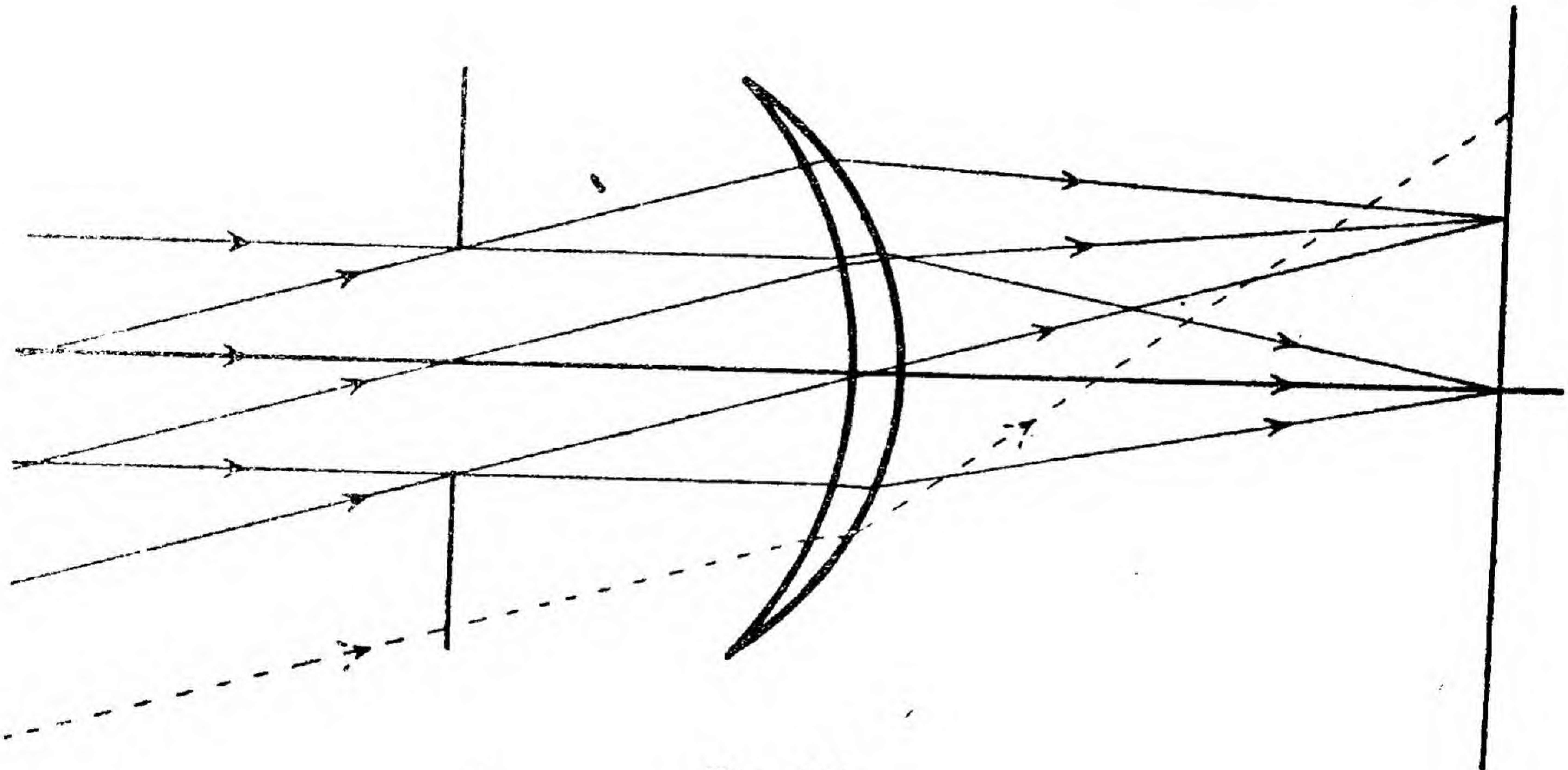


Fig. 111.

stop is arranged so that oblique rays go through that part of the lens which produces equal deviations at the two faces, which reduces the aberrations to a minimum. This also ensures that the field is artificially flattened by giving the tangential and sagittal focal surfaces opposite curvatures, so that the surface of least confusion is flat (Fig. 112). This method of flattening the field increases the astigmatism, so the definition falls off towards the edge of the field. No attempt is made to correct for spherical aberration at the centre of the field, but the loss of definition due to spherical aberration and astigmatism is kept down by restricting the aperture of the lens to f/11. Enlargements are not possible with these cameras. A rather better type of fixed-focus camera has a meniscus combination which is achromatised for two suitable colours, which improves the definition all over the field.

The lenses used in the second class of camera must be designed to

keep down the five aberrations of a thin lens, since these will produce serious effects on the definition and shape of the image at apertures of f/4·5 over a field of 50°. The first step was made by the production of

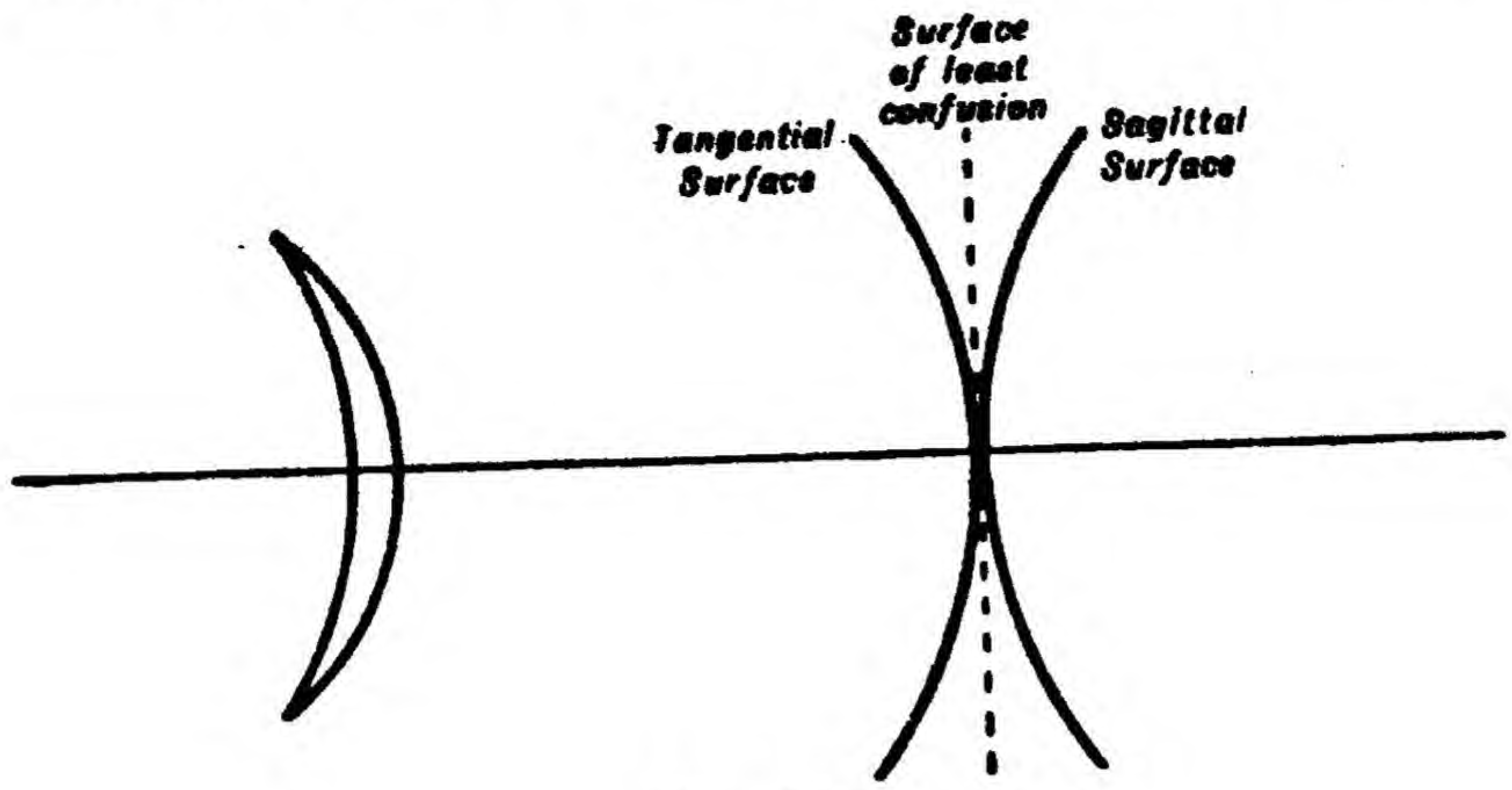


Fig. 112.

the **anastigmat**, which is corrected for spherical aberration, chromatic aberration, astigmatism, and curvature of the field. The spherical aberration is reduced by a suitable bending of the faces of the lens. The conditions for the correction of the remaining three defects are given by the two equations

$$\frac{\omega_1}{f_1}+\frac{\omega_2}{f_2}=0$$

$$n_1 f_1 + n_2 f_2 = 0$$

which can only be satisfied simultaneously, if two glasses are obtained for which $n_1\omega_1=n_2\omega_2$. This was impossible with the ordinary crown and flint glasses, since increased dispersive power is accompanied by increased refractive index (Table 10). This led to a search for new kinds of optical

TABLE 10

Glass.	n for sodium D Light.	ω
Crown.	1·522	0·0168
Flint.	1·624	0·0256
Dense Barium Crown.	1·612	0·0180
Light Flint.	1·579	0·0244

glass and finally resulted in the production of the new Jena glasses. The dense barium crown glass has, as its name dense implies, a large refractive index but the low dispersive power of a crown glass, while the light flint glass has a low refractive index with the high dispersive power of a flint glass. No two glasses have yet been found to satisfy the above relation exactly, but an achromatic lens can be made up of barium crown and light flint glass, which is much freer from astigmatism and curvature

of the field than the old type, and is called a new achromat. A better result has been obtained by a suitable combination of an old and a new achromat with a stop placed in the appropriate position in front of the lens, the residual astigmatism and curvature of the two combinations being arranged to oppose each other (Fig. 113). This is called an anastigmat.

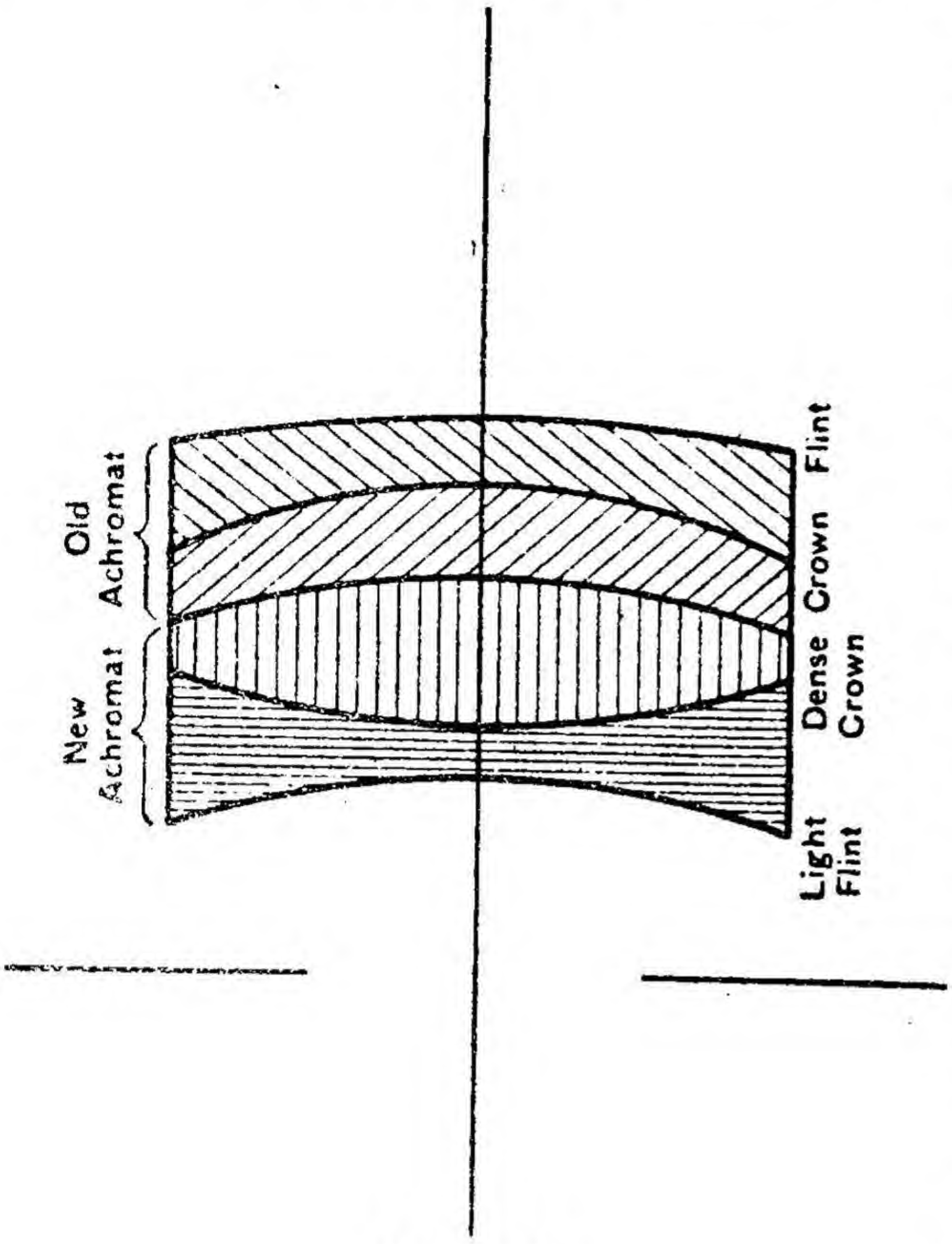


Fig. 113.

To obtain the still greater apertures needed for Press work without any loss of definition and to keep down distortion to a minimum, it is possible to use two anastigmats of the above type placed symmetrically about a stop. The first lens and the stop produce pincushion distortion, while the stop and the second lens produce barrel distortion and the two can be arranged to cancel out one another. It can also be shown that the coma produced by the two lenses is in opposite directions, and so this defect is thereby reduced.

The other classes of lens must be specially designed for the requirements peculiar to their use. It is essential that cameras used for aerial surveys should be free from distortion and so special attention is paid to this point; one way is to decrease the distortion by reducing the field of view of the camera. Lenses used in wide-angle cameras obtain the necessary definition over the large field by working at apertures as small as f/16 and so obtain the desired effect at the expense of long exposure. It really amounts to this: the photographer would like a camera to take photographs over a very wide field with a very short exposure, whose definition is so good that they will allow of a large degree of enlargement. This means that the lens designer must produce a lens combination which will produce a flat image of a large object with good definition over the whole image. It is both theoretically and practically impossible to fulfil all these demands at the same time, and the photographer must make up his mind whether he wants a moderate result in all the above three respects, or whether he will sacrifice one demand in order to obtain greater excellence in the other two respects.

61. THE TELEPHOTO LENS

If a camera, whose lens system is equivalent to a thin lens of focal length 10 cm., is used to take a photograph of a man 2 metres high stand-

ing 4 metres from the lens the image will be formed 10·25 cm. from the lens. This is near enough to the focal plane to assume that the images of all objects further than 4 metres from the lens are formed at the second focal plane of the lens. On this assumption, the length y of the image on the photographic plate of an object subtending an angle 2α at the lens is given by $y=2f\tan\alpha$, where f is the focal length of the lens. If the same camera is now used to photograph the same man standing in the middle of a cricket pitch 160 metres from the camera lens, the value of $\tan\alpha$ is 1/40 its former value and so the length of the image is correspondingly decreased. In fact, the photograph will be too small to be of any use. It can be increased by increasing the focal length of the lens, but this would make the camera itself inconveniently large. Therefore photographs of distant objects are taken with a **telephoto lens**, which is equivalent to a single lens of long focus, but avoids the inconvenience of making the camera long. It consists of a converging lens with a diverging lens placed behind it, so that a ray parallel to the axis of the system passes through the converging lens first and is refracted to the focus F (Fig. 114). This acts as a virtual object for the diverging

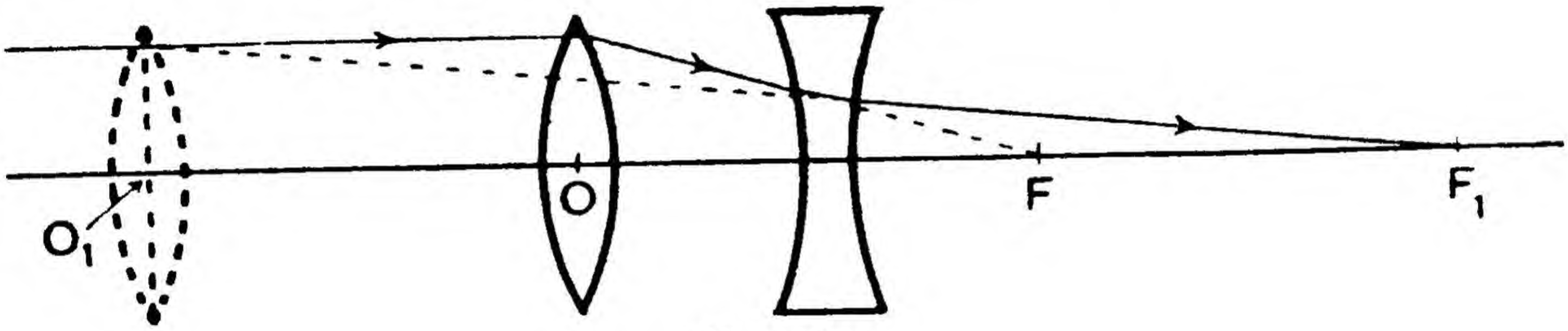


Fig. 114.

lens, which finally causes the ray to cross the axis at F_1, where the photographic plate is put. By producing the final emergent ray backwards to cut the original incident ray, it is seen that the combination is equivalent to a thin lens of focal length O_1F_1 situated at O_1; so the focal length of the combination is considerably greater than the distance OF_1 from the front lens to the photographic plate. In practice the system of lenses used is more complicated than this, since the introduction of the concave lens is so arranged as to magnify the image produced by the convex lens alone about three times, the remaining magnification being done by enlarging the photograph taken with this telephoto lens. Hence it is essential to have good definition, and this means that each component must be suitably corrected for the various lens aberrations.

62. DEPTH OF FOCUS

This account of camera lenses will be concluded with a discussion of the depth of focus of a thin lens, since the results obtained are interesting and throw light on how this essential property of a camera is obtained. It may be mentioned that the results apply to actual camera lenses,

although they are not thin lenses. Let B be the point image of a point object A formed by a convex lens, the numerical value of whose focal length is c (Fig. 115). If a photographic plate is placed at B perpendicular to the axis of the lens, then a point A_1, whose point image is at B_1, is sufficiently well focussed if the circular patch of light formed on the plate by the rays diverging from B_1 has a diameter δ, where $\delta = 0{\cdot}25$ mm. (Art. 60).

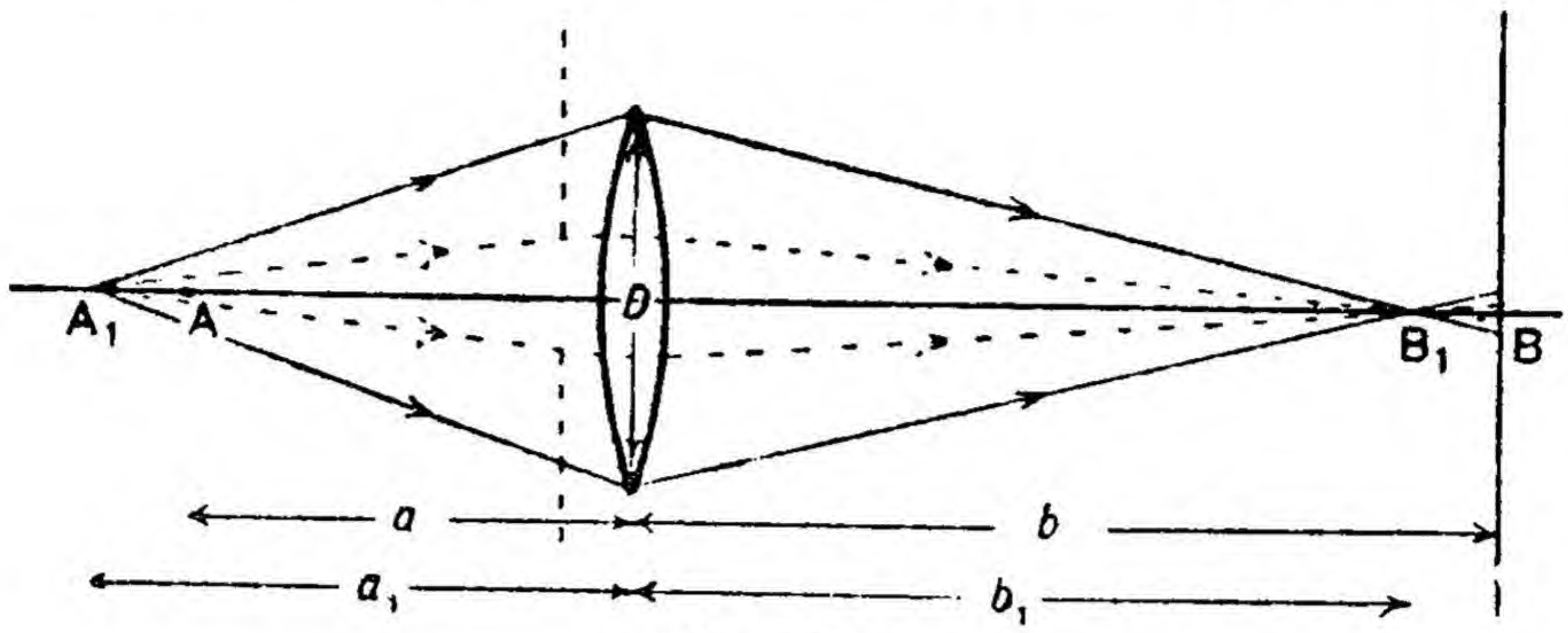


Fig. 115.

Any point further from A than A_1 would not be sufficiently well focussed on the plate and the distance AA_1, which will be denoted by t_1, is the depth of focus of the lens on the side of the object remote from the lens. It is at once evident that, if the effective diameter of the lens is reduced by putting a stop in front of it, as shown in dotted lines, the circular patch on the photographic plate due to the rays diverging from A_1 and

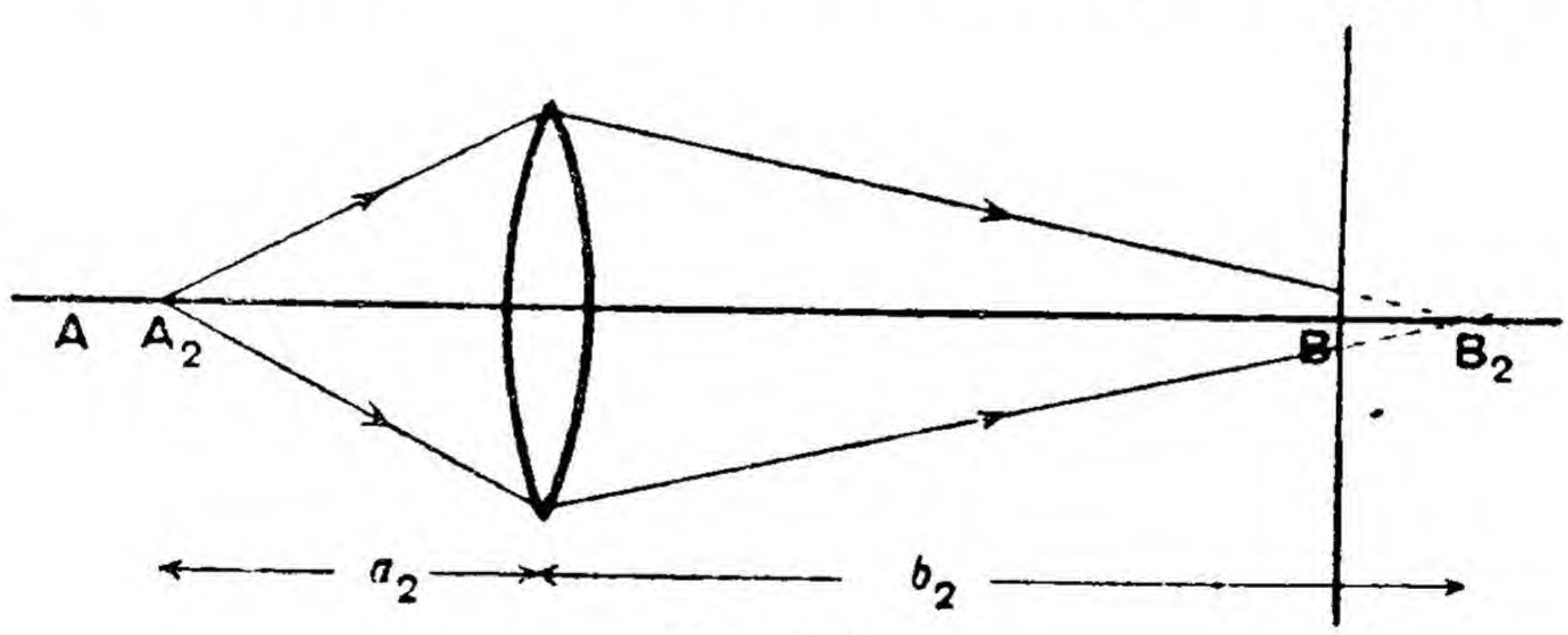


Fig. 116.

passing through the lens will have a smaller diameter than before. These rays are shown in dotted lines in Fig. 115. Therefore a point further from A than A_1 can now be sufficiently well focussed on the plate, and so the depth of focus of a lens increases as the effective diameter of the lens is decreased.

The actual value of the depth of focus can be calculated in this way. From the relation

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

since $f = +c$, $u = -a$, and $v = +b$ for the points A and B

$$\therefore \frac{1}{b} + \frac{1}{a} = \frac{1}{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

Similarly for the points A_1 and B_1

$$\frac{1}{b_1}+\frac{1}{a_1}=\frac{1}{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

If D=the diameter of the lens,

$$\therefore \frac{\delta}{D}=\frac{b-b_1}{b_1}$$

$$\therefore b_1=\frac{b}{1+\frac{\delta}{D}}$$

$$\therefore \frac{1}{b_1}=\frac{1}{b}+\frac{\delta}{bD}$$

Substituting this value of $\frac{1}{b_1}$ in equation (41)

$$\frac{1}{a_1}=\frac{1}{c}-\frac{1}{b}-\frac{\delta}{bD}$$

From equation (40)

$$\frac{1}{c}-\frac{1}{b}=\frac{1}{a}$$

$$\therefore \frac{1}{a_1}=\frac{1}{a}-\frac{\delta}{bD}$$

$$\therefore a_1=\frac{abD}{bD-a\delta}$$

and

$$t_1=a_1-a$$

$$\therefore t_1=\frac{a^2\delta}{bD-a\delta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

In just the same way, for the depth of focus AA_2, denoted by t_2, on the same side of the object as the lens (Fig. 116)

$$\frac{1}{b_2}+\frac{1}{a_2}=\frac{1}{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

$$\frac{\delta}{D}=\frac{b_2-b}{b_2}$$

$$\therefore \frac{1}{b_2}=\frac{1}{b}-\frac{\delta}{bD}$$

whence

$$a_2=\frac{abD}{bD+a\delta}$$

and

$$t_2=a-a_2$$

$$\therefore t_2=\frac{a^2\delta}{bD+a\delta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

So the total depth of focus t is given by

$$t=t_1+t_2$$

$$\therefore\ t=\frac{a^2\delta}{bD-a\delta}+\frac{a^2\delta}{bD+a\delta}\quad . \quad . \quad . \quad (45)$$

A very striking result follows from this relation, in addition to the confirmation of the conclusion, established above, that the depth of focus increases as the diameter of the stop decreases. If $bD=a\delta$, or $D=\frac{a}{b}\delta$, then $t_1=\infty$ and $t_2=\frac{a}{2}$. That is, if any thin lens forms an image of an object, the diameter of the stop can be adjusted so that the depth of focus stretches from half-way between the lens and the object to infinity, a truly amazing result! To consider the case given at the beginning of Art. 61 of a man standing 4 metres from a camera lens of focal length 10 cm.: the plate must be placed 10·25 cm. behind the lens to receive a sharply focussed image. If the diameter of the stop is adjusted to be $\frac{400}{10{\cdot}25}\times 0{\cdot}25=10$ mm., which is $\frac{f}{10}$, the depth of focus will be from 2 metres from the lens away to infinity. A portrait photographer will not work at this aperture, partly because he would want to use a shorter exposure, but also because he wants the background out of focus in order that the person may stand out well in the photograph.

EXAMPLES ON CHAPTER VII

1. The eye can be regarded as a thin lens whose greatest focal length is 2·5 cm. placed 2·5 cm. in front of the retina. Suppose a person can see print distinctly at a distance of 10 cm. from the eye, what must be the focal length of the lens now? Hence find the accommodation of the eye in dioptres, assuming the far point to be at infinity and the near point at 10 cm.

2. A man suffering from presbyopia has his far point at infinity and his near point at 50 cm. from the eye. Calculate the nature and focal length of the lens needed to bring his near point to 25 cm. Where will his far point be when wearing these spectacles and how much will an object at 50 cm. from his eye be magnified by them?

3. A person requires concave lenses of power 2 dioptres to correct his myopia. Where is his far point? If his near point is 15 cm. from the eye, where will it be when wearing his spectacles? Treating the optical system of the eye when the ciliary muscles are relaxed as a thin lens of focal length 2·5 cm., how far must the portion of the retina on the axis of this lens have been displaced to change a normal eye into one with the above degree of myopia? In what ratio is an object at the near point diminished by looking at it through these spectacles?

4. Why does a short-sighted man wear spectacles which consist of concave lenses?

A man who can see an object most distinctly when it is 5 in. from his eyes wishes to read a notice 15 ft. away. What should be the focal length of the lenses in his spectacles? (*Oxford Schol.*)

5. Draw a diagram of the human eye, labelling the parts of optical importance. How does the eye adapt itself (a) to focus objects at different distances, (b) when the illumination varies ?

A short-sighted eye can focus objects situated at distances varying from 15 cm. to 30 cm. What would be its range of vision through a concave lens of three dioptres power placed close to it ?

Sketch the paths of rays from a non-axial point on an object at the new near point through the lens to the receiving surface at the back of the eye. (*N.U.J.B.*)

6. A person wearing spectacles looks over them and is able with one eye to see the moon both directly and also through one of the lenses. The image of the moon is found to be displaced downwards through a distance of one moon diameter. Assuming the moon to subtend an angle of 1° at the eye and the diameter of the spectacle lens to be 3 cm., find the focal length of the lens and explain the defect of vision which it is designed to correct. (*Camb. Schol.*)

7. A short-sighted person has his far point at 50 cm. for vertical lines and at 40 cm. for horizontal lines. From what additional defect is he suffering ? Calculate the focal length of the spectacle lens to cure both defects and suggest suitable powers for each refracting surface.

8. A person suffering from presbyopia with a near point of 40 cm. is found to be myopic as well with his far point at 100 cm. It is decided to fit him with bifocals ; calculate the focal length of each component to give him a far point at infinity and a near point at 20 cm. Calculate his near point when using the upper component and his far point when using the lower one.

9. Discuss the way in which a person of defective eyesight is fitted with spectacles, explaining how the defect is diagnosed and how the correct lenses are prescribed.

10. Discuss the specification of colour and colour vision.

11. A camera lens is required to take photographs of scenes up to a width of 20 yds. on a plate 6 in. square. Find the focal length of the lens, assuming that the scene will be 20 yds. from the lens, and find how far the plate must be from the lens. Give the specification of the lens if it is equi-convex of refractive index 1·60. If the camera is used to photograph a very distant scene, what adjustment must be made and what angle will the scene included on the plate subtend at the camera lens ?

12. A camera lens is to take photographs of objects 6 ft. high at a distance of 10 ft. from the lens on a film 3¼ in. long. Calculate the focal length of the lens required. It is to be a doublet of crown and flint glass, achromatic for the F and H lines, spherical aberration is to be minimised by making the surface facing the plate plane and the surface facing the object convex, and the two lenses are to be cemented together ; calculate the focal length and the radii of curvature of the surfaces of each lens, using the table of refractive indices given in Chapter VI.

13. Discuss the design of camera lenses, showing clearly how the various defects of the image are reduced to a minimum in the various cases and how the design varies with the purpose of the camera.

14. Discuss the main essentials of a good photographic lens. (*Oxford Schol.*)

15. What qualities are desirable in the lens of a camera to be used for taking motion pictures ? (*Oxford Schol.*)

16. Explain the purpose and mode of action of a telephoto lens, illustrating your argument by a numerical example. For what defects is such a lens corrected ?

17. What is the depth of focus of a lens and why is it so important to have a large depth of focus in a landscape camera lens, but a small depth of focus with a portrait lens ? Prove that the depth of focus increases as the diameter of the stop in front of the lens decreases.

Chapter VIII

OPTICAL INSTRUMENTS

63. INTRODUCTION

The theorems concerning the thin lens and the mirror, which have been deduced from the axioms about rays of light, will now be applied to the study of the more complicated optical instruments, such as the telescope, the microscope, and the projection lantern. As with the simple lens and the mirror, these instruments were discovered first and the theorems of Geometrical Optics were used afterwards to explain how they work and to improve their design. The treatment will be divided into two stages; in the first place, the mode of action of the various instruments will be discussed on the assumption that they consist entirely of thin lenses, and then the improvements which have been introduced to transform these "thin lens" instruments into the modern optical instrument will be outlined.

The purpose of all these optical instruments is to make objects appear larger, that is, to produce an image whose **size** is greater than that of the object when seen with the unaided eye. In order to understand how this is done in each case, it is necessary to have clear ideas on the meaning of size. How big an object or image appears to be depends not only on its linear dimensions, such as its length, but also on its distance away from the observer's eye. A reasonable definition of **the size of an object as viewed by a given observer is the ratio of the length of the object to its distance from the eye of the observer.** This is equal to the angle subtended at the eye by the object when its size is so small that the above ratio can be equated to the angle in radians. The **magnifying power** of an optical instrument is then defined as $\dfrac{\text{The Size of the Image}}{\text{The Size of the Object}}$.

The reader should distinguish this quantity from the magnification, which refers to the ratio of the lengths of the image and object.

64. THE MAGNIFYING GLASS

The magnifying glass consisting of a single converging lens is used nowadays by the geologist for a superficial examination of rock specimens and by the jeweller and watchmaker. The perfection of certain specimens of antique Grecian woodwork containing very intricate carved designs suggests

that the Greeks invented the magnifying glass. As its name implies, its purpose is to magnify an object and its mode of action will be understood from Fig. 117, in which the object AA′ is placed inside the first principal focus of the lens, which produces a virtual erect image BB′. This image is viewed by the eye, which is placed close up to the lens. Remembering that the size of an object is equal to the ratio of its length to its distance from the eye, it will be seen that the sizes of the object AA′ and the image BB′ are the same. But, if the lens has a focal length of 2 or 3 cm., it is clear that the eye could not focus the object if the lens were removed, and so the lens produces an image of the object whose size is greater than that which the object can have when viewed by the unaided eye. The magnifying power of this instrument is accordingly defined as

$$\frac{\text{The Size of the Image}}{\text{The Size of the Object When at the Distance of Distinct Vision}}$$

since this is the nearest point at which the object can be seen clearly

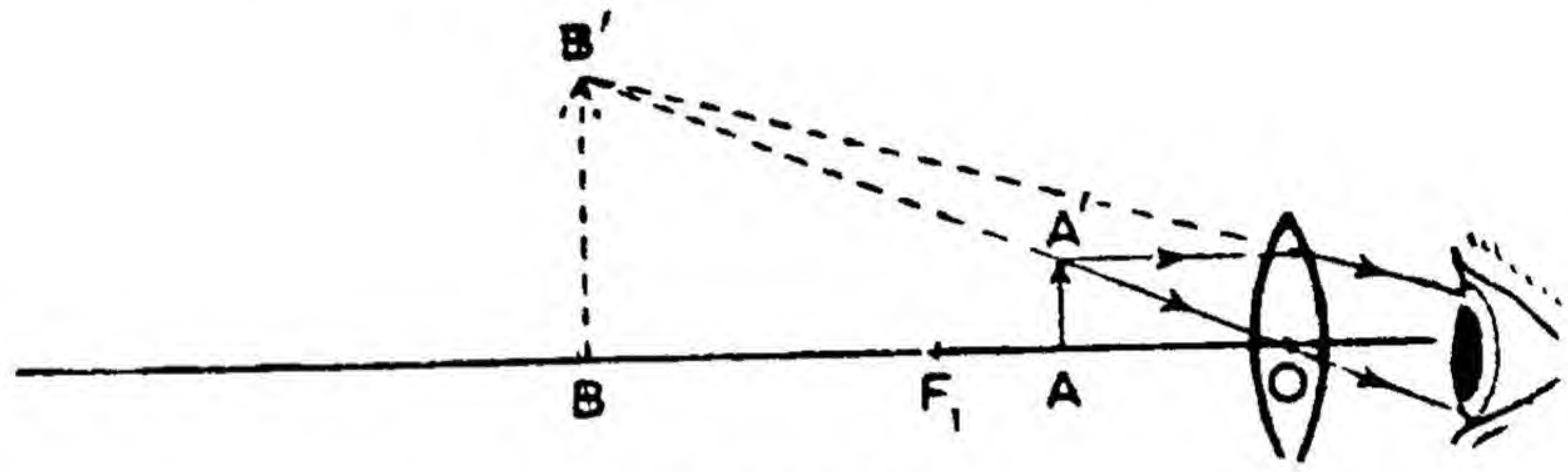


Fig. 117.

without strain with the unaided eye. If the image BB′ is also formed at the same distance, the magnifying power is equal to

$$\frac{\frac{BB'}{D}}{\frac{AA'}{D}}=\frac{BB'}{AA'}$$

where D is the distance of distinct vision. This assumes that the eye is close up to the lens.

Now $\frac{BB'}{AA'}=\frac{OB}{OA}=\frac{b}{a}$, where b=the distance OB, a=the distance OA.

But
$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

and $v=-b$, $u=-a$, $f=+d$, the numerical value of the focal length of the lens.

$$\therefore \frac{1}{-b}-\frac{1}{-a}=\frac{1}{+d}$$

$$\therefore \frac{1}{a}-\frac{1}{b}=\frac{1}{d}$$

$$\therefore \frac{b}{a}=1+\frac{b}{d}$$

$$\therefore \text{Magnifying power}=1+\frac{b}{d}$$

But
$$b=\text{D}$$

$$\therefore \text{Magnifying Power}=1+\frac{\text{D}}{d}$$

So the less the focal length of the lens, the greater its magnifying power, which is why a short-focus lens with very curved faces is popularly known as a strong lens.

It is much less strain on the eye, if it has to look at objects for a long time, to place the magnifying glass so that the final image is at infinity, since the eye is then unaccommodated. In this case, the object is at the first principal focus of the lens, and so the size of the image is $\frac{\text{AA}'}{\text{OF}_1}$ and the magnifying power is

$$\frac{\frac{\text{AA}'}{\text{OF}_1}}{\frac{\text{AA}'}{\text{D}}}=\frac{\text{D}}{d}$$

The magnifying power is a little less in this case, but the difference is negligible if a lens of focal length about 2 cm. is used, since D is about 25 cm. for a normal-sighted person. The above relation also shows that the magnifying glass is less use to a myopic person than to one whose sight is normal, because the value of D is less for the short-sighted than for the normal-sighted person. This illustrates another way of regarding the action of the magnifying glass. We have seen that the size of the image BB′ is the same as the size of the object AA′, if each is viewed by an eye at O (Fig. 117). But an eye at O could not focus the object AA′, since it is inside its near point. So the function of the magnifying glass is to enable the eye to see distinctly objects inside its near point, while keeping their size the same. That is, it makes the eye artificially short-sighted, and so it is evident that a myopic eye cannot receive so much help from a magnifying glass as a normal eye.

It is natural to expect chromatic aberration to spoil the definition of the image seen in a magnifying glass and to produce colouration, but it is surprising that magnifying powers of 14 can be obtained with very little trouble in this respect. A little examination shows the reason for this freedom from chromatic aberration. If $(\text{BB}')_r$ and $(\text{BB}')_v$ are the lengths of the red and blue image respectively and b_r and b_v their respective distances from the lens,

$$\frac{(\text{BB}')_r}{b_r}=\frac{\text{AA}'}{a} \text{ and } \frac{(\text{BB}')_v}{b_v}=\frac{\text{AA}'}{a}$$

If the eye is close up to the lens, each image subtends the same angle at the eye and so the different coloured images exactly fit on to one another and produce a white image. The different distances from the eye at which the images are formed are exactly compensated by the difference in their lengths. The reader must consult Hardy and Perrin's "Principles of Optics" to see how the magnifying glass has been improved so as to give either greater field of view at a low magnifying power or greater magnifying power at the expense of a smaller field of view.

65. "THIN LENS" OPTICAL INSTRUMENTS: THE ASTRONOMICAL TELESCOPE

We shall now consider in terms of the thin lens the mode of operation of the telescope and microscope, both of which magnify objects. But the telescope magnifies objects whose size is small because of their great distance away from the observer, while the microscope is designed to magnify objects which owe their small size to their small linear dimensions. It is evident already that the two cases can change continuously from

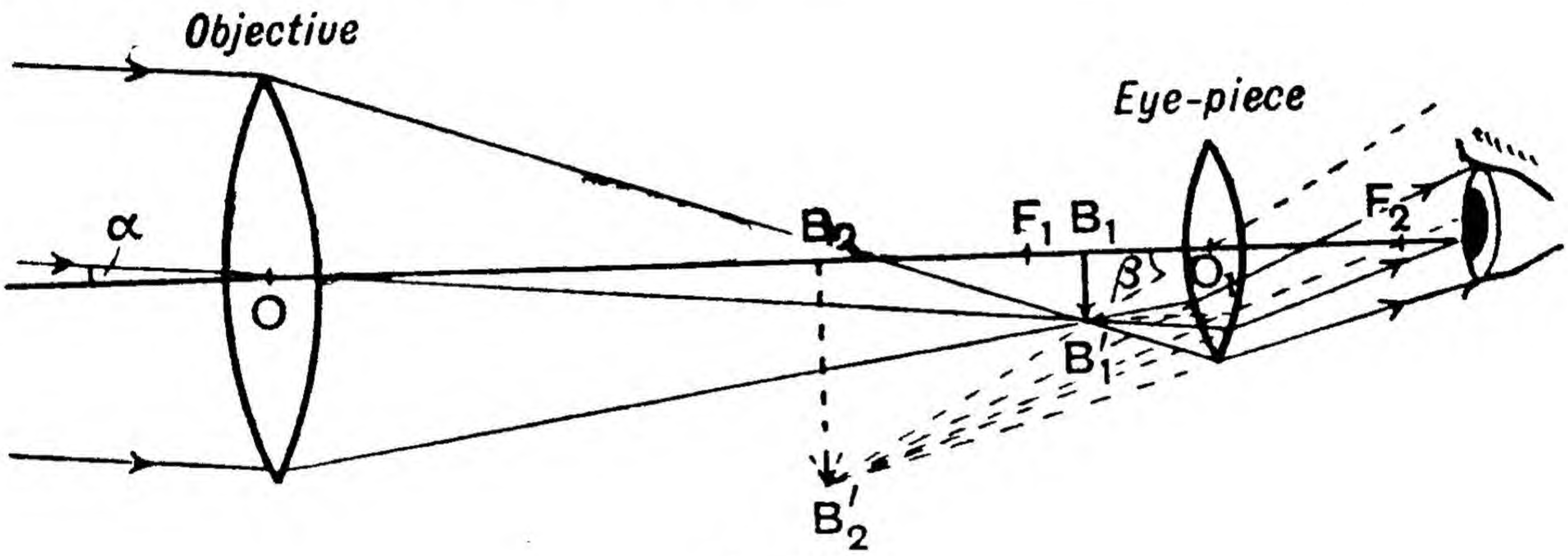


Fig. 118.

one to the other, and it will be seen that the two instruments are essentially the same in principle. We shall commence with the astronomical telescope, which was first designed by Kepler in 1611 and is so called because it is used in astronomy and any other type of observation, where the fact that the final image is inverted is no handicap. The telescope consists of two lenses (Fig. 118), the **objective**, a long focus converging lens which is pointed at the object, and the **eyepiece**, a short focus converging lens which is put to the eye. If the common axis of these two lenses is pointed at the bottom of the object, the diagram shows the passage of three rays from the top of the object through the instrument. Since the object is at a great distance from the objective, the three rays from the top of the object enter the objective almost parallel to one another, but inclined at a finite angle α to the axis of the telescope. This angle incidentally is equal to the size of the object, which subtends the same angle at the eye as it does at the objective owing to the length of

the telescope being so small compared with the distance away of the object. The objective forms a real inverted image B_1B_1' of the object in its second focal plane and inside F_1, the first focus of the eyepiece; this image is then magnified by the eyepiece acting as a magnifying glass, the final image being formed at B_2B_2'. The exact position of B_2' is fixed by drawing two "construction rays" from B_1' shown in dotted lines, one $B_1'O_1$ passing through the centre of the lens and the other parallel to the axis of the lens passing through its second principal focus F_2. B_2' is at the point where these rays meet when produced backwards, and the paths of the three rays from the top of the object after emerging from the eyepiece can then be drawn by making them appear to diverge from B_2'. It will be seen that the magnifying is done in two stages. First the image B_1B_1' is produced, and, if the eye is placed at the distance of distinct vision from it, its size is $\frac{B_1B_1'}{D}$, while its size to the unaided eye is α or $\frac{B_1B_1'}{d_o}$, where d_o is the focal length of the objective. So the magnifying power of the objective is $\frac{d_o}{D}$. The focal length of the refracting telescope at the Yerkes Observatory is 65 feet, and so this objective has a magnifying power of $\frac{65 \times 12}{10}$ or 78. Then the magnifying power of the eyepiece of focal length d_e is $\frac{D}{d_e}$, if the final image is at infinity as is most comfortable. So the magnifying power of the whole telescope is

$$\frac{d_o}{D} \times \frac{D}{d_e} = \frac{d_o}{d_e}$$

This relation only applies when both the object and the final image are at infinity, and a rather more general expression can be obtained in the following way. Let the object be at a distance a_1 from the objective and let its length be l. If the image B_1B_1' is formed at a distance b_1 from the objective, the length of the image is $\frac{b_1}{a_1}l$. If this image is a distance a_2 from the eyepiece and the final image B_2B_2' is formed at a distance b_2 from the eyepiece, then the length of this final image is $\frac{b_1b_2}{a_1a_2}l$. Therefore the size of the object is $\frac{l}{a_1}$, assuming as before that the length of the telescope is negligible compared to the distance of the object from the eye, and the size of the image is $\frac{b_1l}{a_1a_2}$, assuming that the eye is close up to the eyepiece. Hence the magnifying power of the telescope is $\frac{b_1}{a_2}$.

This applies to any positions of the object and the final image. If the object is at infinity, $b_1 = d_o$, and if the final image is at infinity also, $a_2 = d_e$ and so the magnifying power is $\frac{d_o}{d_e}$, which is the same as we obtained above. It is now evident why the telescope consists of a long focus objective and a short focus eyepiece, since this combination has the greatest magnifying power.

A telescope is not only required to produce a highly magnified image; it should also produce a bright image and cover a wide field of view. These two requirements will now be discussed in turn. In considering the brightness of the image, it is necessary to distinguish between extended objects such as planets and point objects such as stars. If a telescope forms an image at infinity of an extended object at infinity, the rays from the point of the object at which the telescope is pointed will go through it as shown in Fig. 119. It is evident that the emergent pencil is narrower

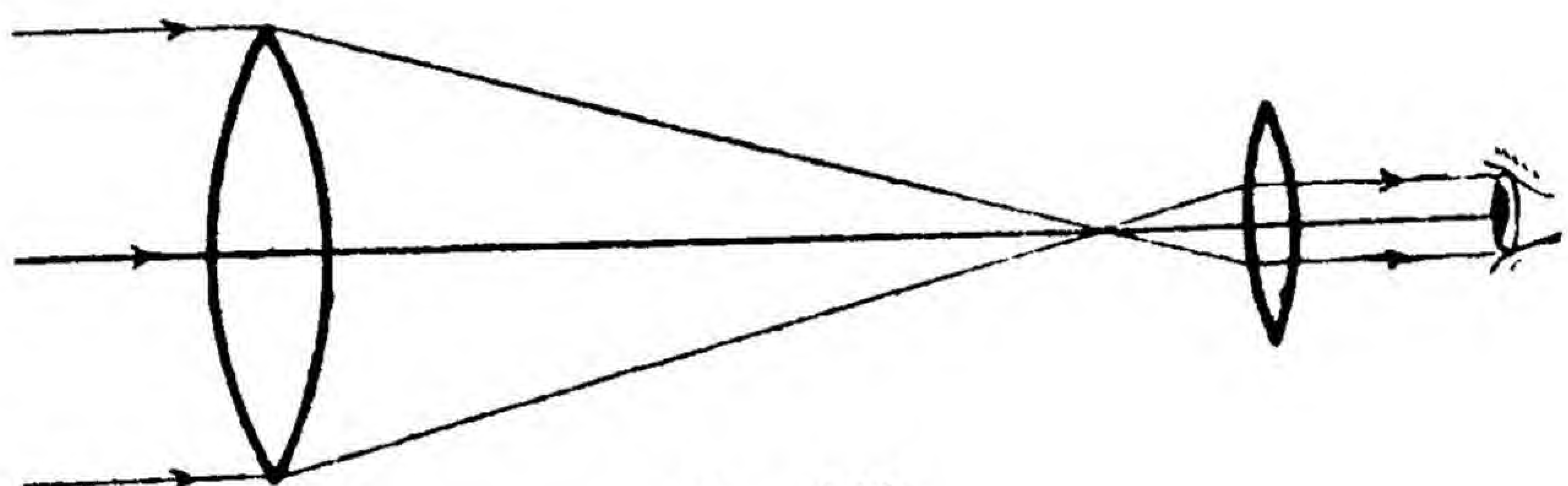


Fig. 119.

than the incident one in the ratio of the focal lengths of the objective and eyepiece, since the pencil comes to a focus at the common focus of the two lenses. Thus diameter of emergent pencil $= \frac{\text{diameter of incident pencil}}{\text{magnifying power}}$.

It is evident that, if the emergent pencil does not fill the pupil of the eye, the final image is not as bright as it might be, and so it is necessary to increase the diameter of the objective until this is the case. Any further increase beyond this is useless, since it merely means that the whole of the objective is not being utilised. But it can be shown that the final image is now no brighter than the object appears to the unaided eye. If the magnifying power is m and the telescope is designed so that the emergent pencil just fills the pupil of the eye, then m^2 times more light enters the eye from each point of the object, when the telescope is used, than when the object is observed with the unaided eye. But the area of the image is m^2 that of the object, and so the brightness of both is just the same. So a telescope can be designed to give an image as bright as, or fainter than, the object, but never brighter. In practice the image is usually fainter than the object, since other considerations enter into the fixing of the diameter of the objective, as will be seen later (Art. 143).

If the telescope is used to form an image of a point object such as one of the fixed stars, the image will be brighter than

the object seen by the unaided eye. For the brightness of the background will be the same in the two cases, but, since the objective of the telescope collects more light from the star than the pupil of the eye in the ratio of their respective diameters squared, the image seen in the telescope will be brighter than that seen with the unaided eye. For this reason telescopes with objectives of large diameter are especially valuable in the investigation of very distant stars, and, if the diameter of the objective of a telescope is doubled, the distance of the faintest star visible through it is doubled.

The **field of view** of a telescope in the object space is defined as the angle subtended at the eye of the longest object, the whole of which can

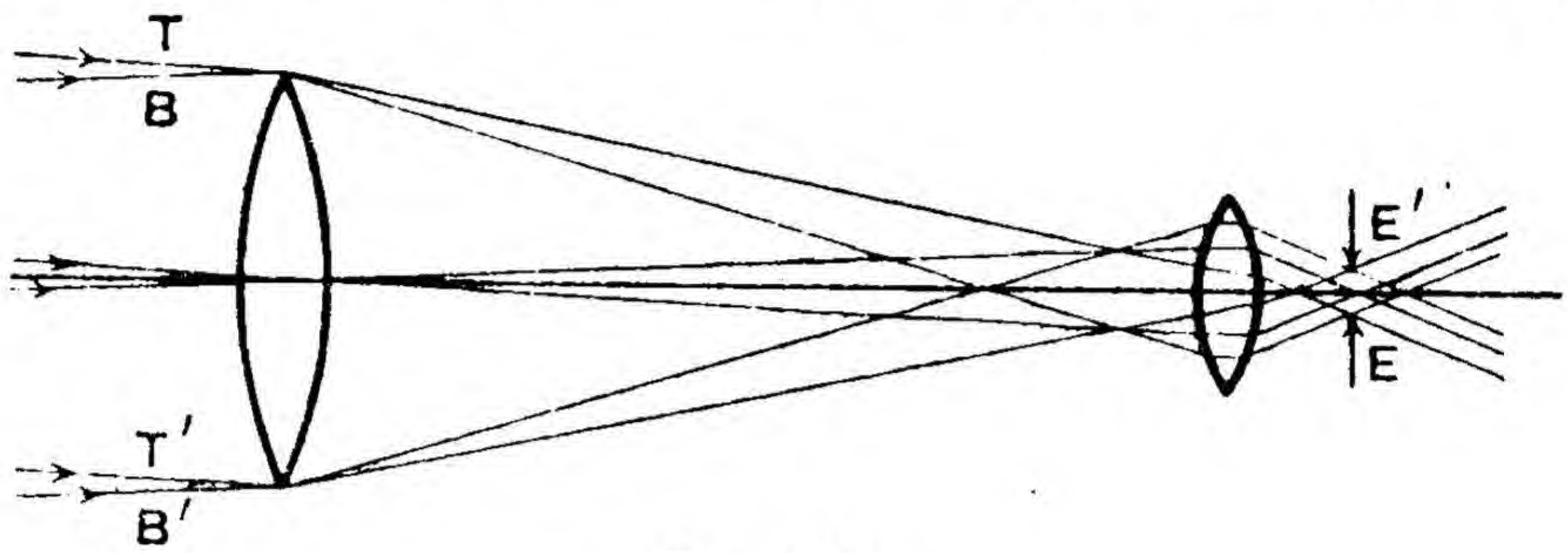


Fig. 120.

just be seen; the angle subtended at the eye by the image of this object is called the field of view in the image space. In order to see what factors control the field of view, let us examine Fig. 120, which represents the passage of rays from the top and bottom of an object at infinity through the telescope which is adjusted to produce the final image at infinity. It is easily seen that the field of view is a maximum when the eye is placed at EE′, which represents the circle formed by the intersection of

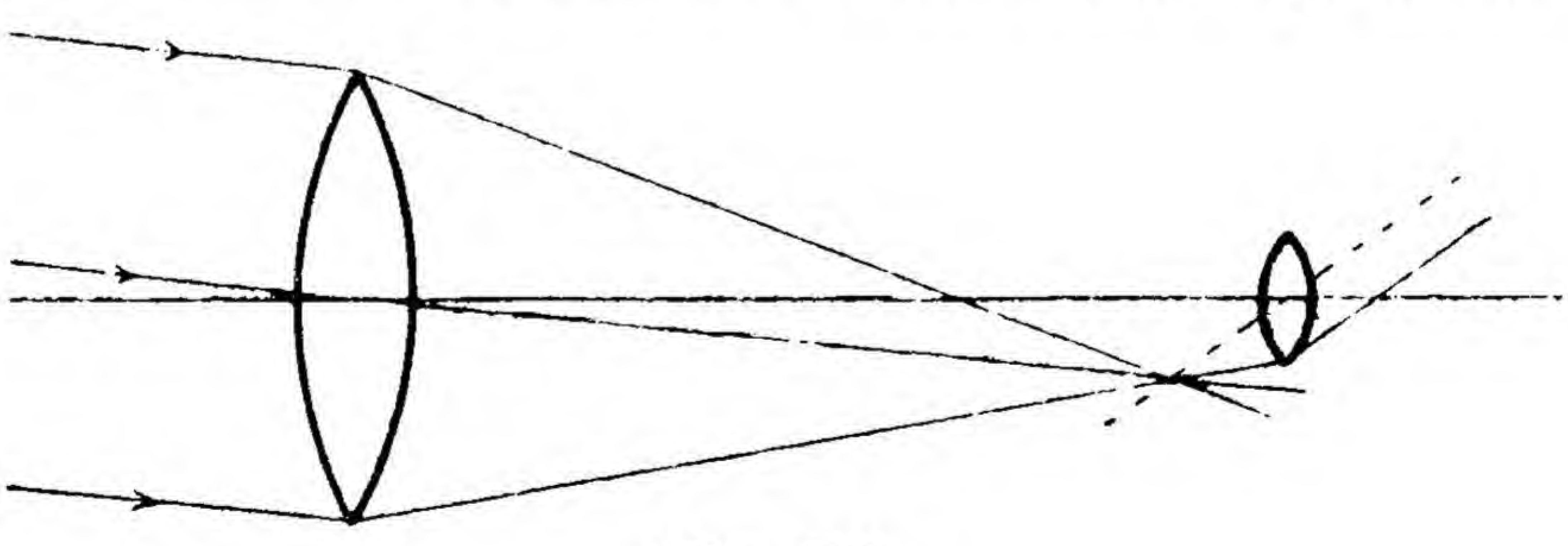


Fig. 121.

the two pencils from the top and bottom of the object. This circle, called the **eye-ring**, is the image of the objective formed by the eyepiece, because the rays T′ and B′ which converge on to the bottom of the objective are brought to a focus at E′ by the eyepiece and similarly T and B converging on to the top of the objective are brought to a focus at E. The eye-ring is just outside the focus of the eyepiece and its diameter is equal to that of the objective divided by the magnifying power of the telescope, since it is equal to the diameter of the emergent pencil shown in Fig. 119. It is usual to place a cap over the eyepiece of a telescope with a hole in it, such that it coincides with the eye-ring. Consequently,

when the eye is put to this hole in the cap, the field of view is a maximum. With the eye in this position it can be seen that the field of view is largely controlled by the diameter of the eyepiece, and Fig. 121 shows the rays from a point on the very edge of the field passing through the telescope. It will be seen that only one ray from this point finally emerges from the eyepiece, and so the illumination will fall off badly towards the edge of the field. To avoid this, a stop is inserted at the focal plane of the objective as shown in Fig. 122, so that the central ray from a point on the edge

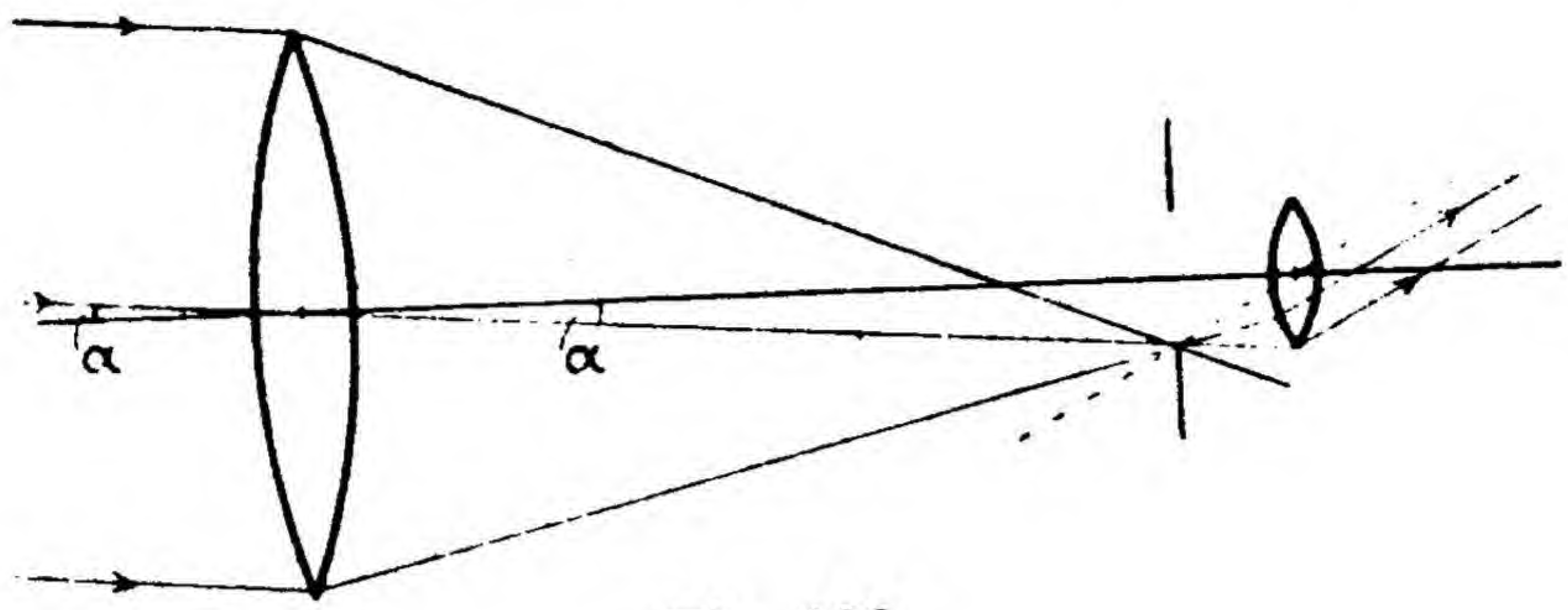


Fig. 122.

of the field can pass through the eyepiece. Under these conditions, the field of view in the object space is 2α, which is $\frac{M}{d_o+d_e}$, where M is the diameter of the eyepiece. The field of view of a simple telescope consisting of an objective of focal length 50 cm. and an eyepiece of focal length 10 cm. and diameter 6·5 cm. is $\frac{6 \cdot 5}{60}$, or 6·2°, which is 3·1° on each side of the axis. The field of view evidently decreases as the magnifying power increases; and this decrease cannot be compensated for indefinitely by a corresponding increase in the diameter of the eyepiece, since the value of this is limited by the aberrations of oblique pencils. But the reader should contrast the above value of the field of view of a telescope of moderate magnifying power with the 50° which is provided in the cheapest cameras.

The design of an astronomical telescope proceeds therefore along the following lines. The magnifying power to be produced is settled first of all; this fixes the diameter of the objective, which is given by the product of the magnifying power and the diameter of the pupil of the eye. The relative aperture of the objective is then decided by settling how much aberration can be tolerated in the image produced by the objective. If the aperture is to be 1/15, then the focal length of the objective is its diameter × 15 and that of the eyepiece is at once obtained from the magnifying power. The diameter of the eyepiece is then made as large as possible so as to give the biggest possible field of view, but the factor which controls the biggest value for the diameter is the amount of aberration which can be allowed in the oblique pencils. The focal lengths and diameters of the objective and eyepiece have been settled and the

telescope can now be constructed. It will be noted that the radii of curvature of the faces of the two lenses have not yet been fixed, and so that opportunity is left for the correction of both chromatic and spherical aberration. These matters will be dealt with later (Art. 71).

66. THE GALILEAN TELESCOPE

This was the first telescope to be constructed and, since it produces an erect image, it is especially suitable for viewing terrestrial objects. It was probably first made by Lippershey, a Dutch spectacle maker, in

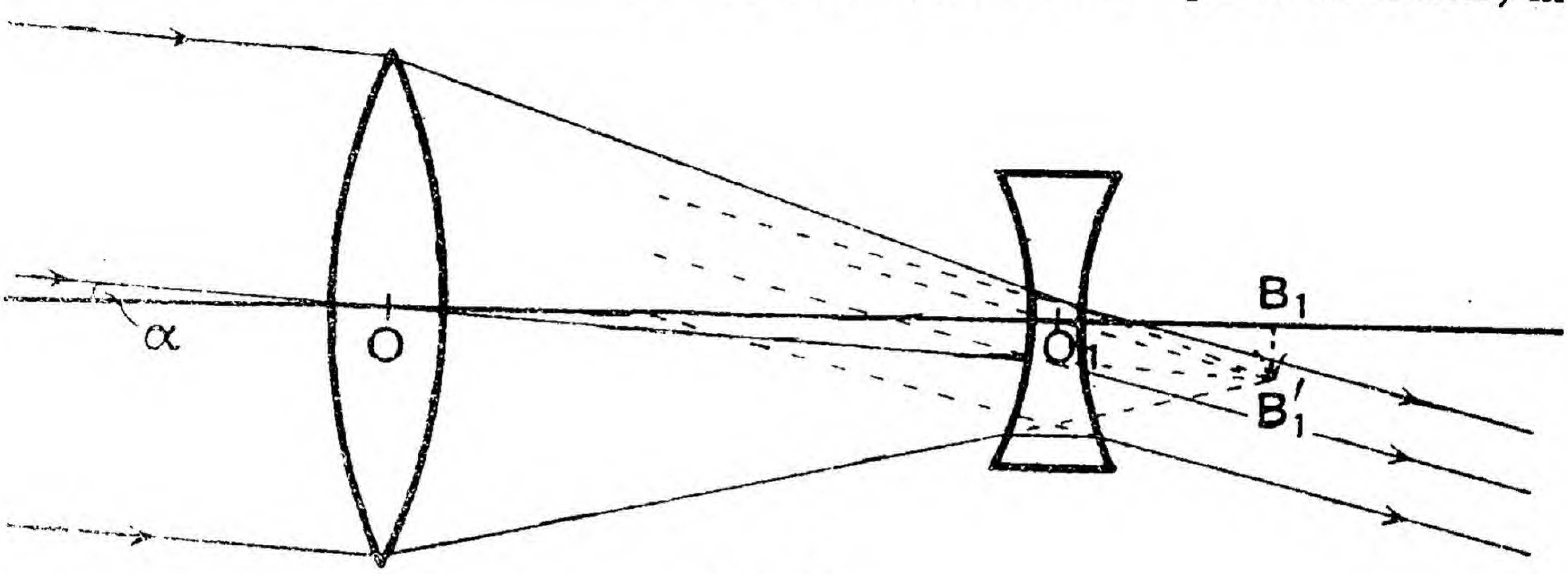


Fig. 123.

1609, and it was from his use of it that Galileo first heard of the instrument and made one himself independently in 1610. It consists of a long focus converging objective together with a short focus diverging eyepiece (Fig. 123). The objective forms a real inverted image B_1B_1' of the distant object in its second focal plane; this acts as a virtual object for the diverging eyepiece, whose first focus coincides with the second focus of the

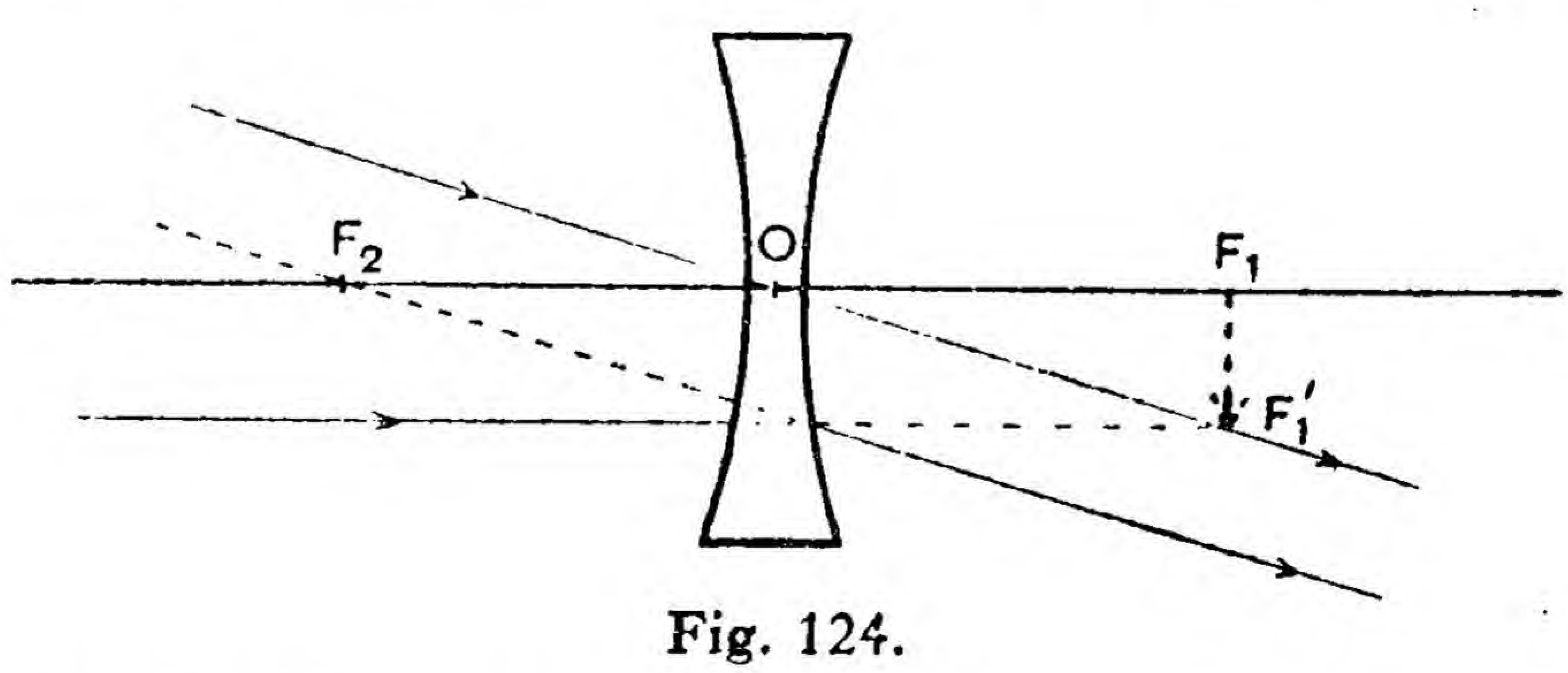


Fig. 124.

objective if the final image is to be at infinity, as is the case in Fig. 123. If a virtual object is situated in the focal plane of a diverging lens, it is shown in Fig. 124 that the rays emerge from the lens as a parallel beam; F_1' is the virtual object and its image is constructed in the usual way and is at infinity behind the lens. Consequently the rays converging to B_1' in Fig. 123 emerge as beam parallel to the construction ray O_1B_1', and so the final image of B_1' is at infinity behind the lens and the image

of the whole object is evidently erect. The telescope magnifies since the angle subtended by the object at the eye is α, while that subtended by the final image is β, if the eye is placed close up to the eyepiece.

The magnifying power is calculated in just the same way as with the astronomical telescope. If the object has a length l and is at a distance a_1 from the objective, the image B_1B_1' formed at a distance b_1 from the objective has a length $\frac{b_1}{a_1}l$. If the eyepiece is at a distance a_2 from this image and treats it as a virtual object, it forms a virtual image at a distance b_2 from the lens. The length of this image is $\frac{b_2b_1}{a_2a_1}l$ and it subtends an angle $\frac{b_1 l}{a_2a_1}$ at an eye close to the eyepiece. The angle subtended by the object at the objective, and so at the eye, is $\frac{l}{a_1}$ and therefore the magnifying power of the telescope is $\frac{b_1}{a_2}$. If it is adjusted so as to produce a final image at infinity of an object at infinity, then $a_2=d_e$ and $b_1=d_o$, and so the magnifying power is $\frac{d_o}{d_e}$, which is the same expression as for the astronomical telescope. It may be noted that this telescope is shorter than the astronomical telescope, since the distance between the lenses

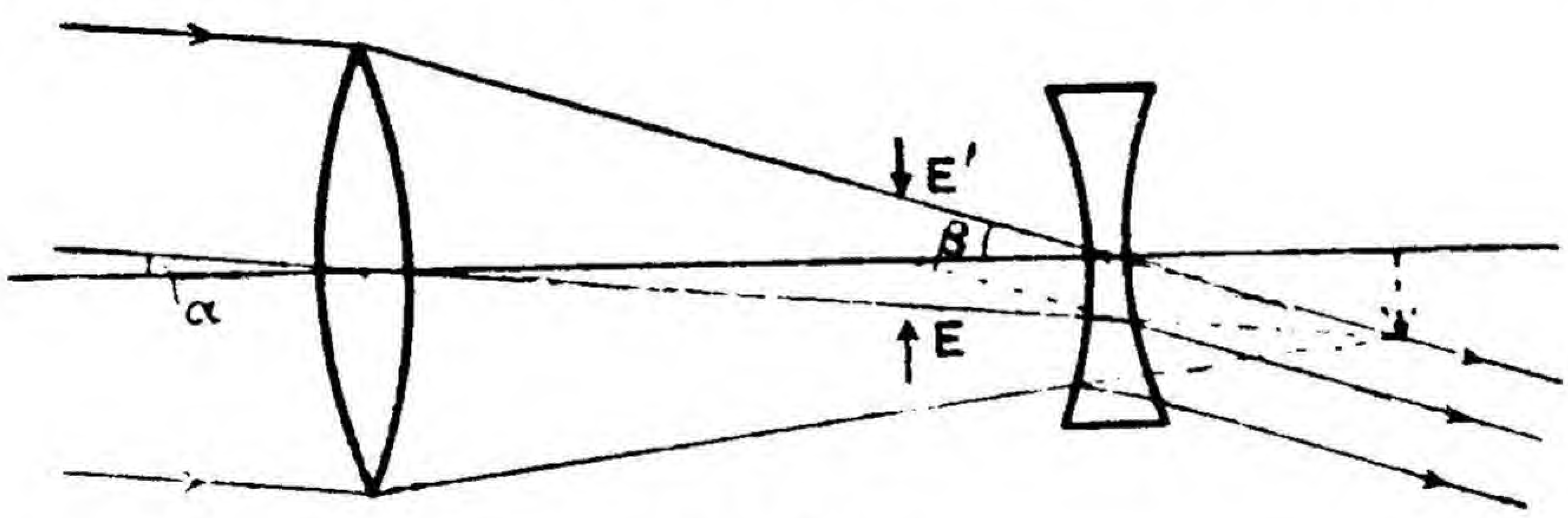


Fig. 125.

is equal to the *difference*, instead of the sum, of the numerical value of their focal lengths.

When considering the field of view of the astronomical telescope, it was seen that the eye must be placed at the eye-ring to obtain the biggest field. The eye-ring is the image in the eyepiece of the objective, and in the case of the Galilean telescope this will be a virtual image situated at EE′ (Fig. 125). It is impossible to get the eye up to the eye-ring, and so the field of view of the Galilean telescope is much inferior to that of the astronomical telescope. Fig. 125 shows the passage of three rays through the telescope from a point on the edge of the field to an eye right up to the eyepiece; the criterion, which decides the edge of the field, is that the ray from the top of the objective shall just pass through the middle of the eyepiece and so enter the eye. The field of view in

the image space is 2β, which is equal to $\frac{M}{d_o - d_e}$, where M is the diameter of the *objective.* In order to obtain a sharp image, it is not possible to make the diameter of the objective greater than $\frac{d_o}{15}$ and so the field of view is $\frac{1}{15}$, neglecting the focal length of the eyepiece. This is only 3·8°. The field of view in the object space is $\frac{3{\cdot}8^\circ}{m}$, where m is the magnifying power of the telescope, and so the Galilean telescope cannot be used at large magnifying powers since its field of view is so small. It is used quite commonly to-day in the form of opera glasses, which are simply a pair of Galilean telescopes mounted parallel to one another; their magnifying power is usually two.

67. THE TERRESTRIAL TELESCOPE

Owing to the small field of view of the Galilean telescope, it was thought desirable to modify the astronomical telescope so as to preserve its large field of view but to produce an erect image. This is done by inserting a converging erecting lens between the objective and eyepiece at a distance of 2f from the real inverted image B_1B_1' of the object produced by the objective (Fig. 126). This lens produces an image B_2B_2' of the same size

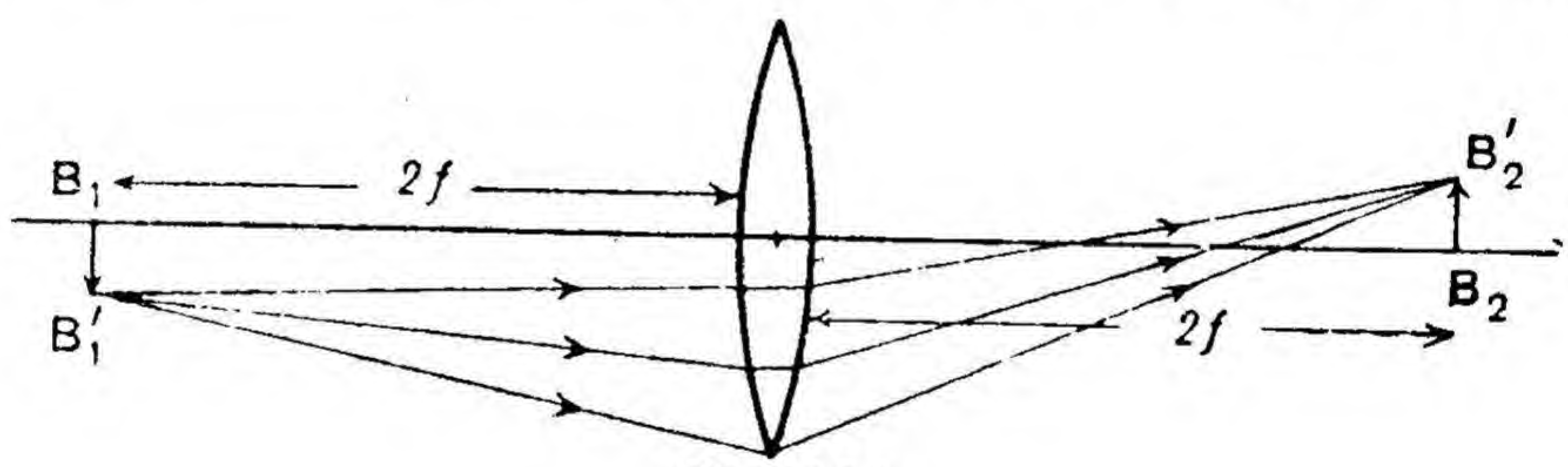


Fig. 126.

as B_1B_1' but the other way up, and this image is magnified by the eyepiece in the usual way. Since the image B_2B_2' presented to the eyepiece is the same way up as the object, so is the final image. The insertion of this erecting lens has the effect of increasing the length of the telescope by four times the focal length of the erecting lens and makes the telescope unduly long. It is therefore necessary to mount it on a tripod to obtain good observations.

68. PRISMATIC BINOCULARS

A successful attempt to combine the erect image and shortness of the Galilean telescope with the large field and magnifying power of the astronomical telescope was made with the production of prismatic binoculars. They are simply a pair of astronomical telescopes, in which the real inverted image produced by the objective is erected by two right-angled isosceles prisms before it is presented to the eyepiece for its final magnifica-

tion. The way in which the image is erected will be clear from Figs. 127 and 128. An examination of the first of these two diagrams shows that, if a ray falls at any angle on an isosceles right-angled prism and suffers two refractions with an internal reflection in between, it emerges parallel to its original direction. The second diagram shows the central rays of pencils from the top and bottom of an object AA′ passing through the objective and on to the prism, and makes it clear that the prism erects the image in a plane perpendicular to the refracting edge. Therefore, in order to erect the image completely, two prisms are needed with their refracting edges at right angles to each other. If the reader obtains a pair of prismatic binoculars and holds them to his eye to look at an object on the same horizontal level as his eye, the objectives are further apart and above the eyepieces. After passing through the objective, the light strikes a prism with its refracting edge vertical, which inverts the image in a horizontal plane and sends the rays inwards and away from his eye; they then strike a prism with its refracting edge horizontal, which inverts the rays in a vertical plane and sends the rays downwards and towards the eyepiece. The complete path of the rays is shown in Fig. 129, and it is seen that the use of the prisms shortens the length of the telescope considerably. It is quite common to have binoculars with a magnifying power of 8 and a field of view in the object space of 6°.

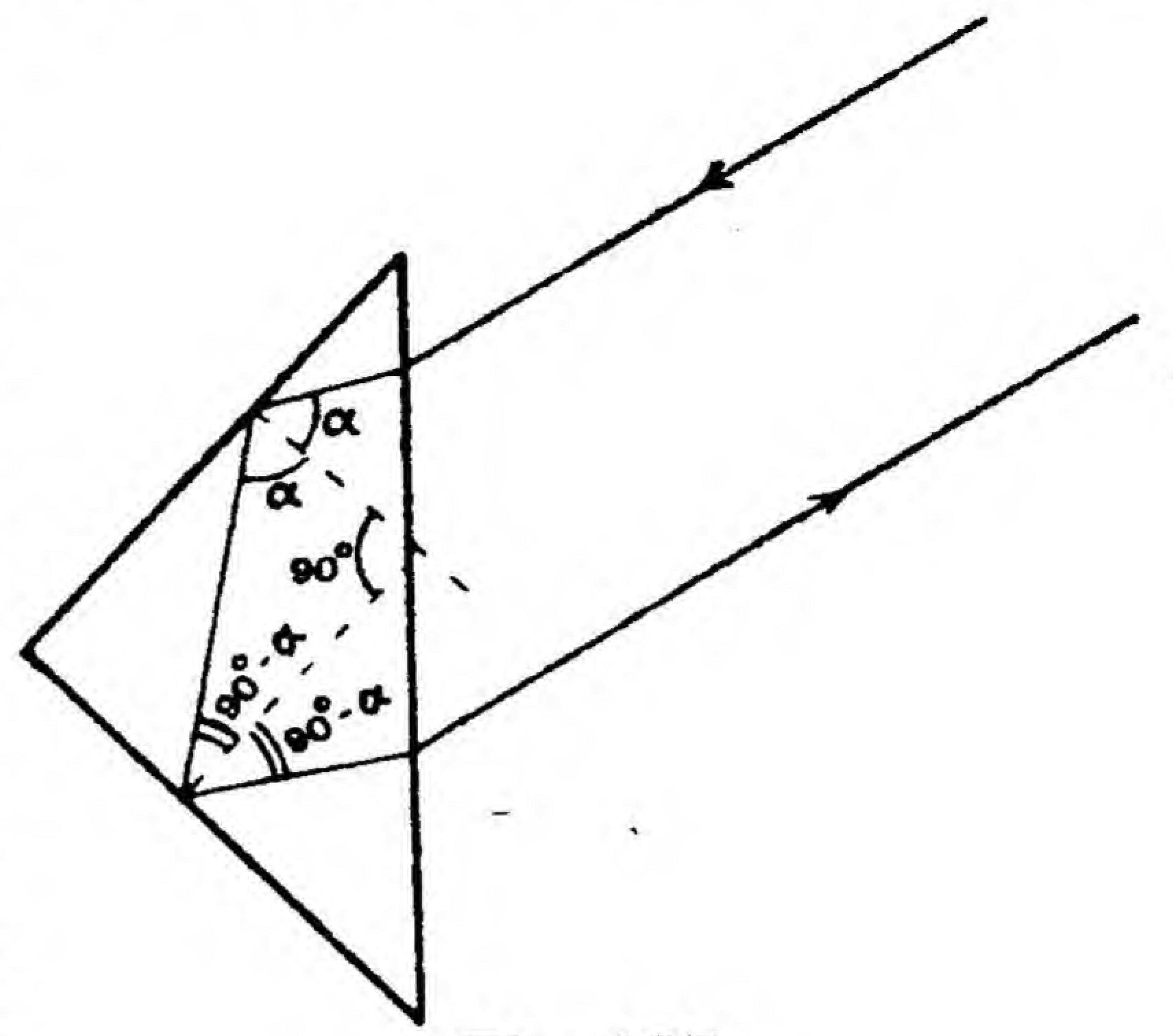


Fig. 127.

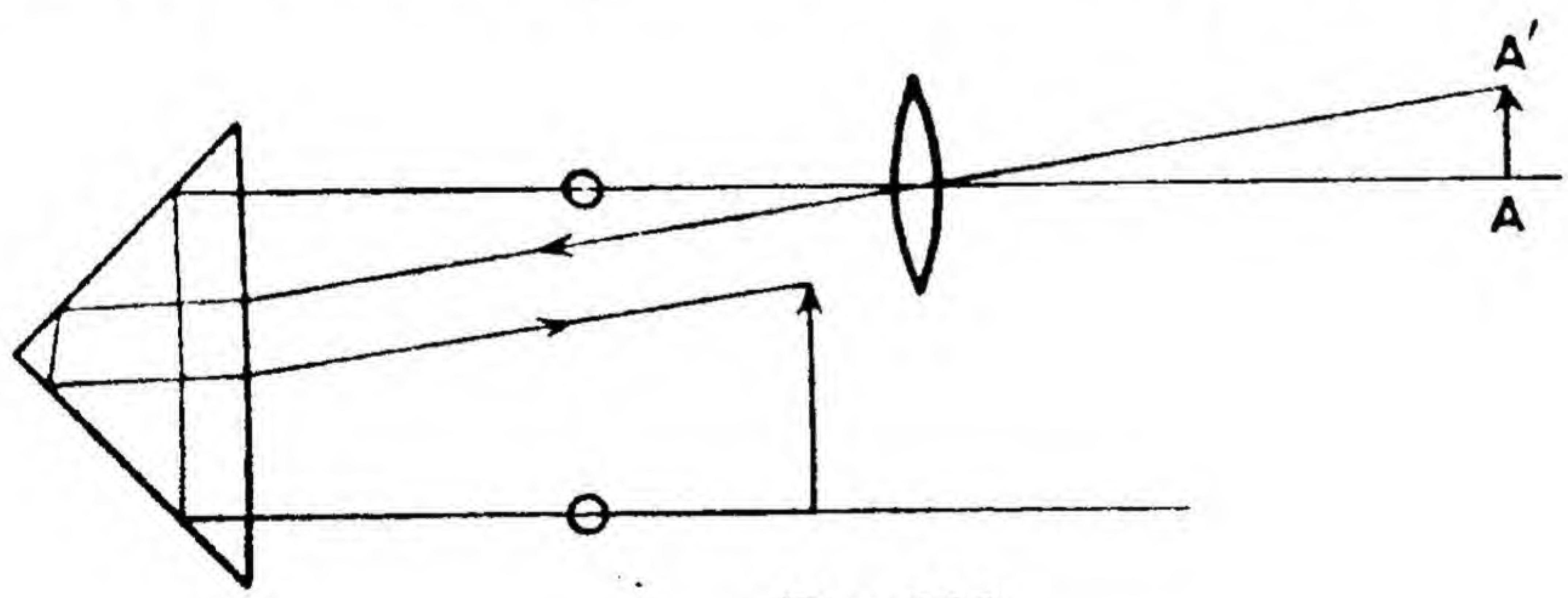


Fig. 128.

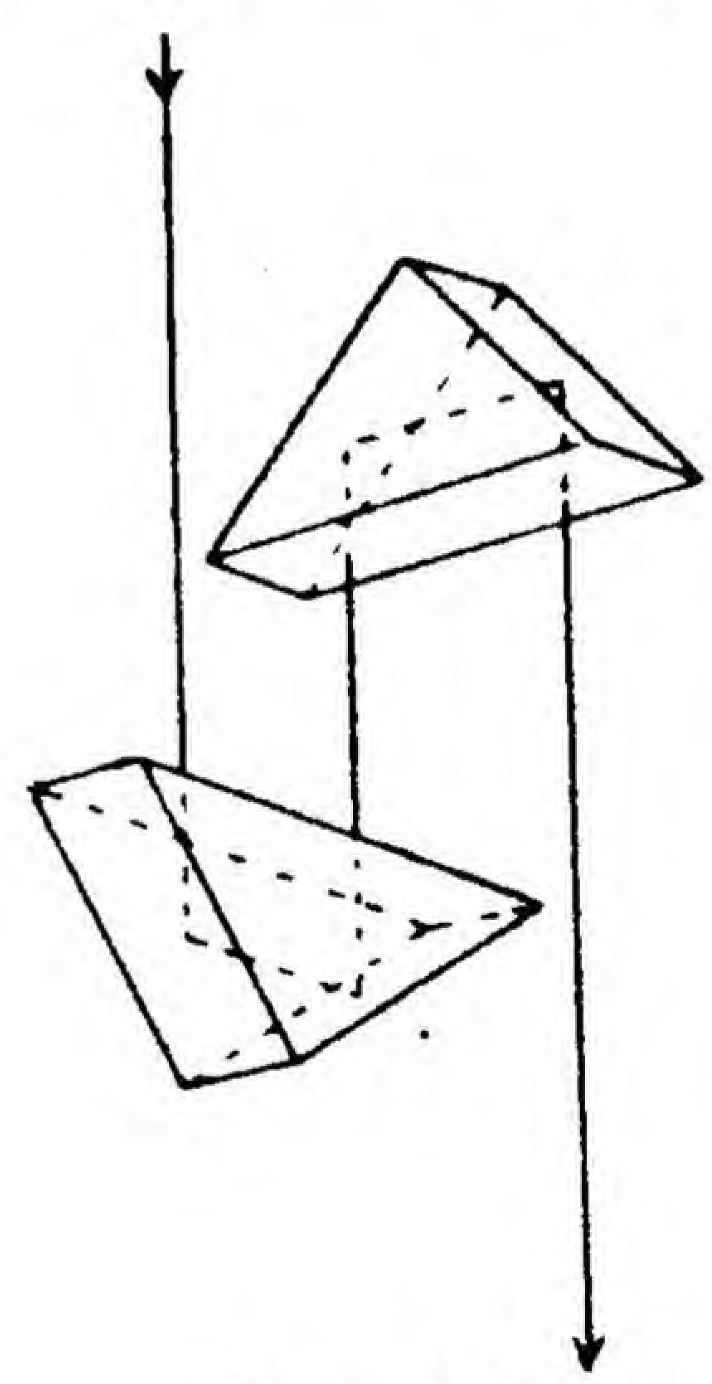


Fig. 129.

69. THE MEASUREMENT OF THE MAGNIFYING POWER OF TELESCOPES

There are three ways of measuring the magnifying power of telescopes. In the first method the telescope is focussed on to a distant blackboard, on which some parallel equidistant lines have been ruled some 5 to 10 cm.

apart. These lines are viewed with one eye through the telescope and with the other eye unaided. After a little practice it will be possible to see both images simultaneously and to place the head so that they are superimposed. Then the number of scale divisions seen with the naked eye contained in one scale division seen through the telescope is counted, and this is the magnifying power of the telescope.

In the second method the whole of the diameter of the objective is illuminated by pointing the telescope at the sky and the eye-ring is focussed on a screen placed at right angles to the axis of the telescope. The cap over the eyepiece is removed for this purpose, if necessary. The diameter of both the objective and the eye-ring are measured, and their ratio is the magnifying power of the telescope.

The above two methods can be used with practical telescopes, but the following method only applies to the simple astronomical telescope. It consists in measuring the focal lengths of the objective and eyepiece by one of the standard methods and finding their ratio, which is the magnifying power.

70. THE REFLECTING TELESCOPE

We have already seen in Chapter 4 how Newton commenced his experiments on the passage of white light through a prism in order to clear up the problem of the blurred and coloured images formed in the refracting telescopes of his day and how he came to the conclusion that this was due to the chromatic aberration of the objective. He also concluded that it was impossible to make an achromatic lens, and so he turned his attention to the reflecting telescope, in which the objective is a concave

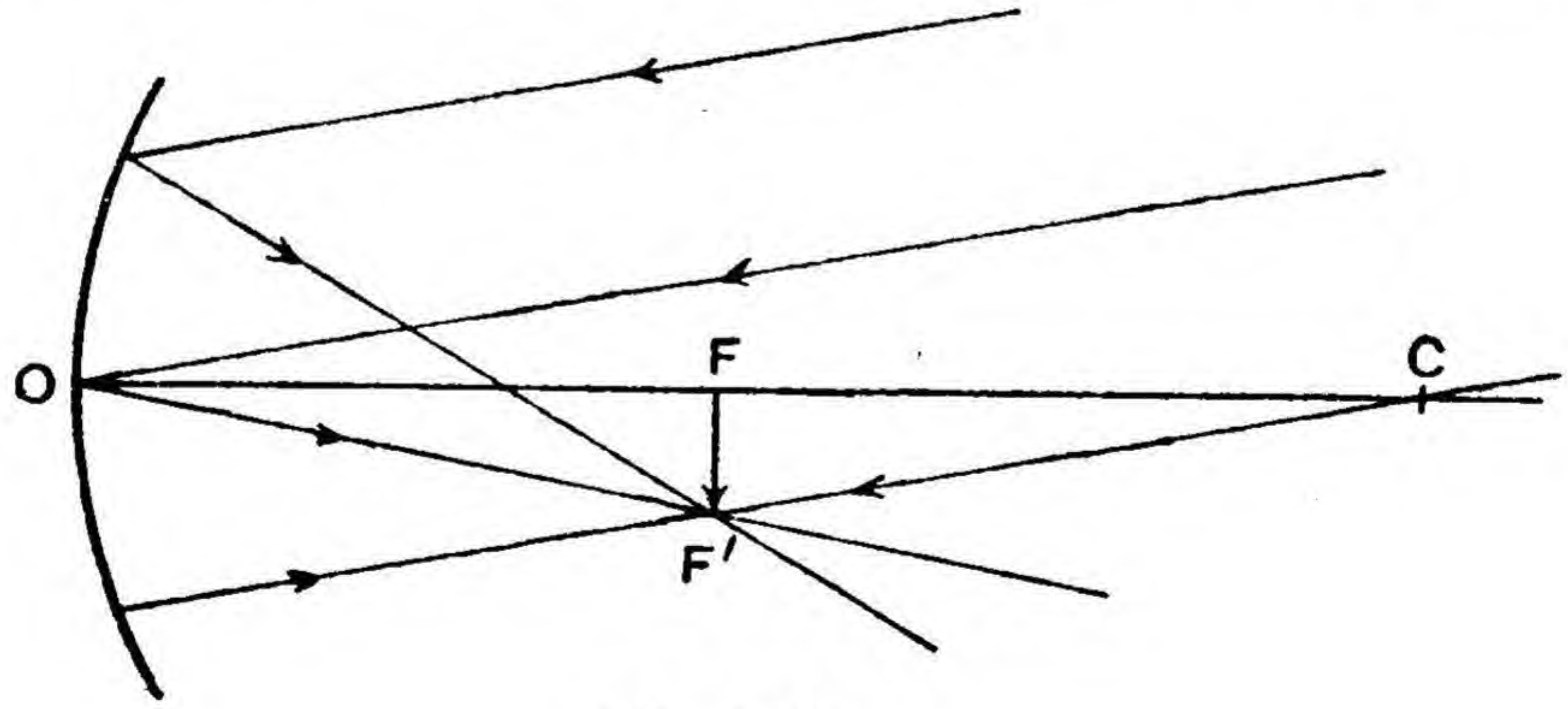
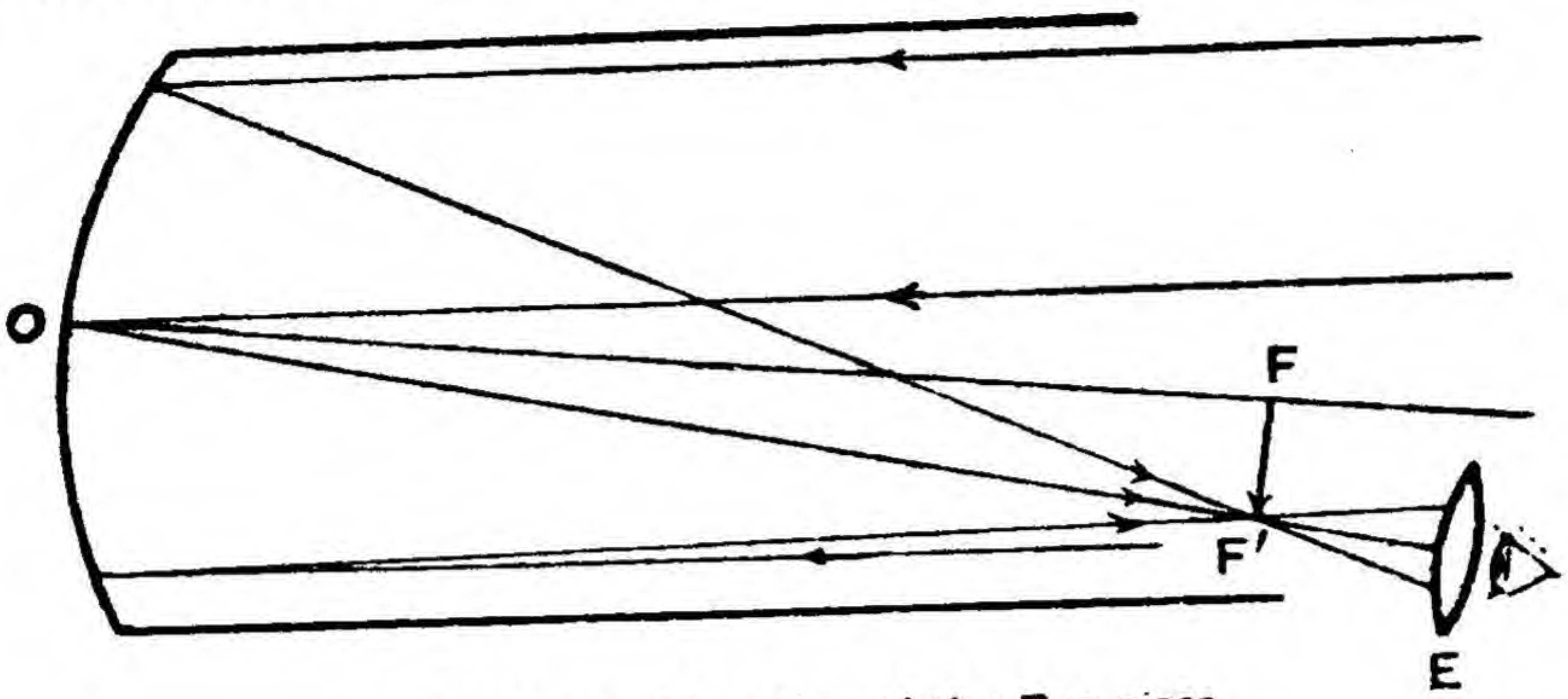


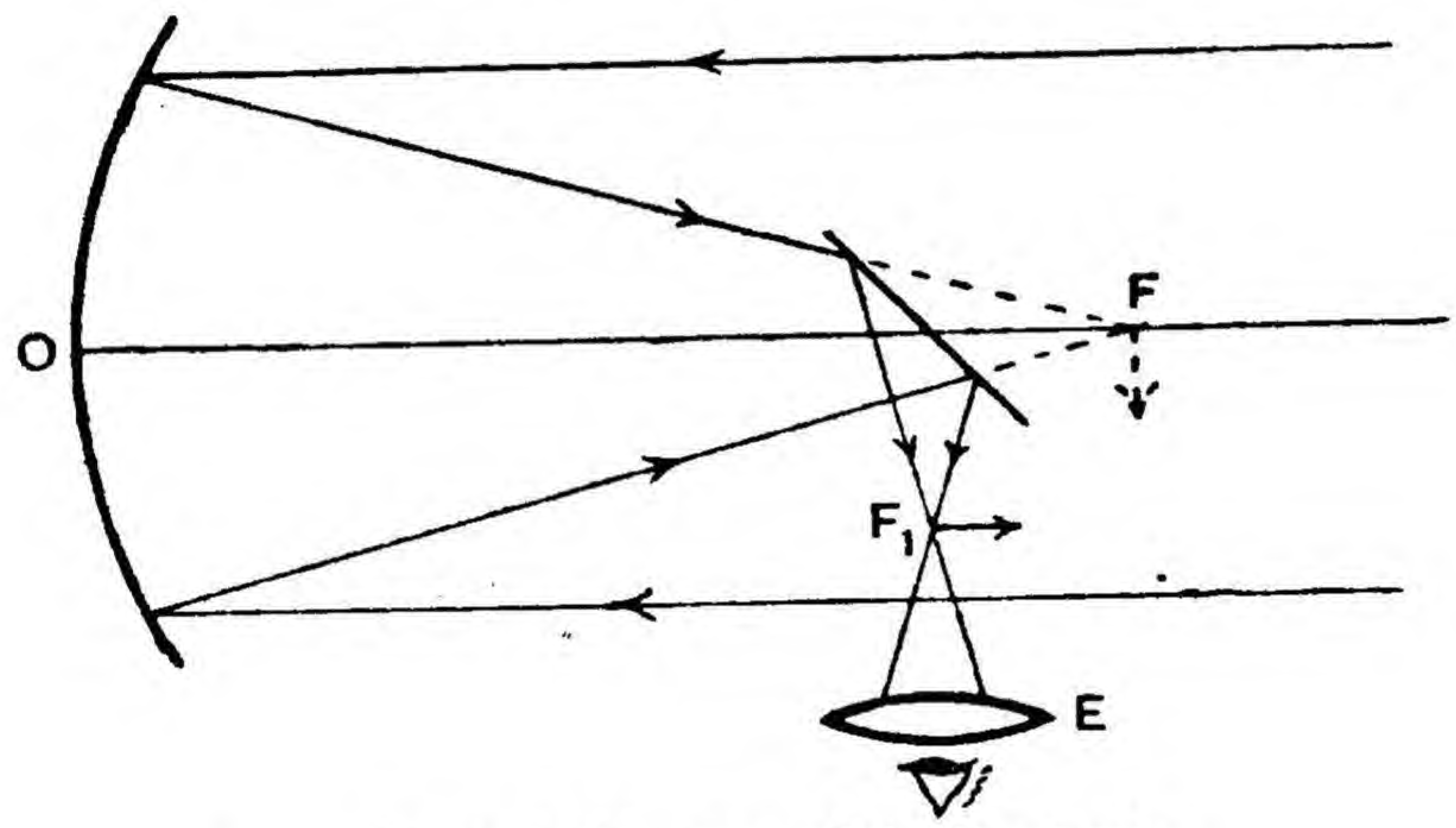
Fig. 130.

mirror of long focus and which is therefore free from chromatic aberration. The principle of the reflecting telescope is just the same as that of the astronomical telescope, except that the real inverted image of the distant object formed by the long focus converging objective is now produced by a long focus concave mirror (Fig. 130). The mirror is a paraboloid of revolution in order to eliminate spherical aberration, for such a mirror brings all rays parallel to the axis to a point focus on the axis, however far from the axis they may be before striking the mirror. If the axis of

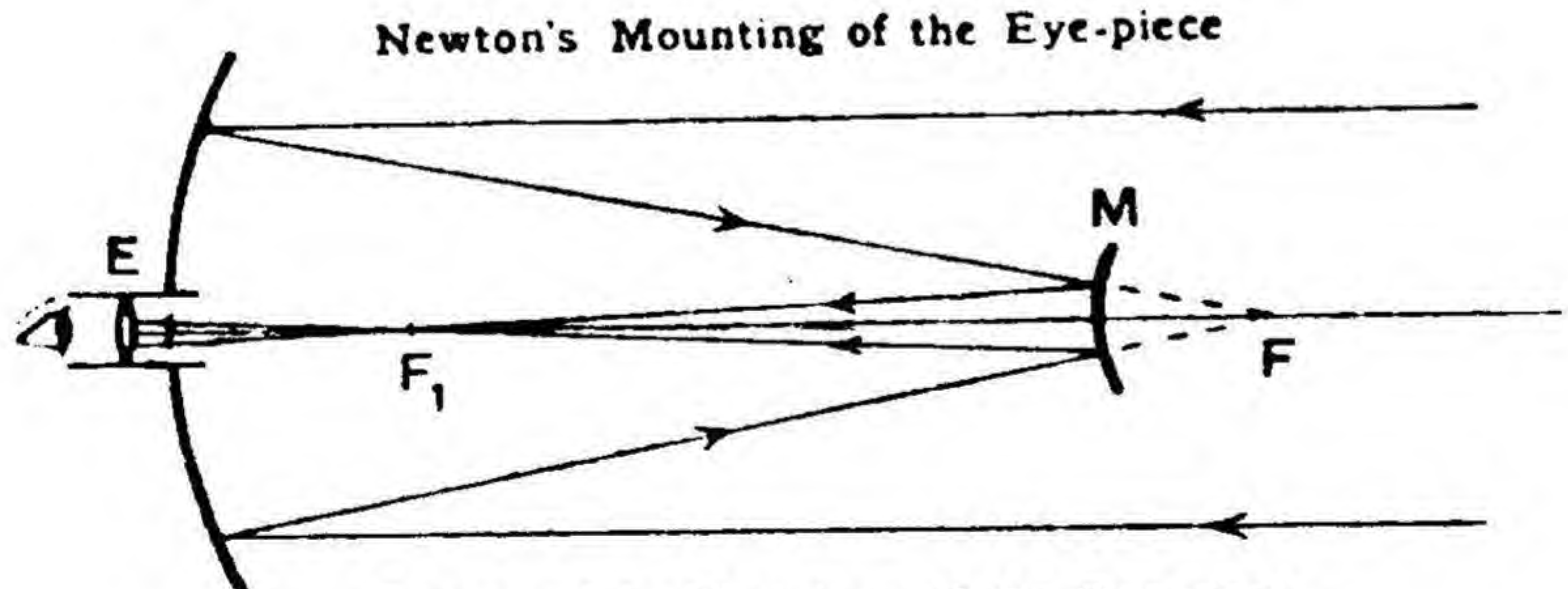
the mirror is pointed at the bottom of the object, then three rays from the top are shown striking the objective in Fig. 130. They come to a focus at the point F′ in the focal plane where it is intersected by the ray passing through the centre of curvature C, and so a real inverted image FF′ of the distant object is formed by the mirror. This image is then examined in an eyepiece consisting of a short-focus converging lens in the usual way.



Herschel's Mounting of the Eye-piece



Newton's Mounting of the Eye-piece



Cassegrain's Mounting of the Eye-piece

Fig. 131.

It is evident that, if the eyepiece is mounted to the right of FF′, the head of the observer will prevent some of the rays from the object from reaching the mirror and so decrease the brightness of the image to a serious extent in the case of mirrors of small diameter. In order to avoid this, various ways of mounting the eyepiece have been introduced from time to time, and the three most common ones are illustrated in Fig. 131. In Herschel's method, the axis OF of the objective is tilted relative to the geometrical axis of the tube in which it is mounted, and so, if the tube is pointed at the top of the object, the image FF′ of the whole object

is formed near to the edge of the tube and so can be examined by the eyepiece E with the minimum of obstruction from the observer's head. In Newton's method, a plane mirror or a right-angled prism is mounted at 45° to the axis just inside the focus F, and so the parallel beam from one point of the object shown in the diagram is turned towards the side of the tube and brought to a focus at F_1, instead of at F. So an image of the whole object is formed at F_1 and is then examined through an eyepiece E mounted in the tube, and the only obstruction of the rays from the object is provided by the small plane mirror or prism, which can be made narrow by arranging to mount it near to the focus F of the mirror. Finally, in Cassegrain's method, a convex mirror M is mounted just in front of the focus F of the paraboloidal mirror. If the telescope is pointed at the bottom of the object, the figure shows two rays from that point striking the paraboloidal mirror, which sends them towards its focus F. Before they reach it they strike the convex mirror M, which finally brings them to a focus at F_1, which is the place where the real inverted image of the whole object is formed. This image is then examined in the eyepiece E.

The reflecting telescope clearly possesses three advantages over the refracting telescope. It is essentially free from chromatic aberration; it is free from spherical aberration for beams parallel to its axis and so for points of the image at, and close to, the focus of the mirror. This means that the definition at the centre of the field is excellent, although it falls off quite rapidly as we go away from the centre, as coma and astigmatism soon become large. Finally, it is practically possible to make mirrors of considerably larger diameter than lenses and so brighter images of stars can be obtained with reflecting than with refracting telescopes. In other words, the range of a reflecting telescope can be made greater than that of the refracting telescope. Thus the telescopes of largest diameter are reflectors. For example, the largest refractor is at the Yerkes Observatory and has an objective of diameter 42 in. and focal length 65 ft., while the largest reflector is at Mount Wilson and has a mirror of diameter 100 in. And a reflector of 200 in. is at present being constructed for use at Mount Palomar. These mirrors are made of glass, or better still quartz, and are covered with a film of silver, or aluminium, to produce a surface of high reflecting power. The majority of astronomical telescopes are refractors, because it is not usually necessary to have objectives of such large diameters as even 42 in., and for technical reasons outside the scope of this book the large refractor is an easier telescope to mount and handle than the large reflector.

71. THE MICROSCOPE

The telescope is used to magnify objects which owe their small size to their large distance from the observer, but the microscope is employed to magnify objects which owe their small size to their small linear dimen-

sions and which are too small to be seen distinctly enough, even when they are at the distance of distinct vision. It consists of an objective made of a very short focus converging lens and an eyepiece consisting of a short-focus converging lens (Fig. 132). The object AA′ to be examined is placed just outside the first focus F_1 of the objective, which forms a real, inverted, magnified image B_1B_1' of it just inside the first focus F_3 of the eyepiece. This is used to magnify the image B_1B_1' like a magnifying glass, the final image being formed at B_2B_2'. It is usual to adjust the position of the eyepiece so that this final image is at the distance of distinct vision, since the eye will automatically accommodate itself for

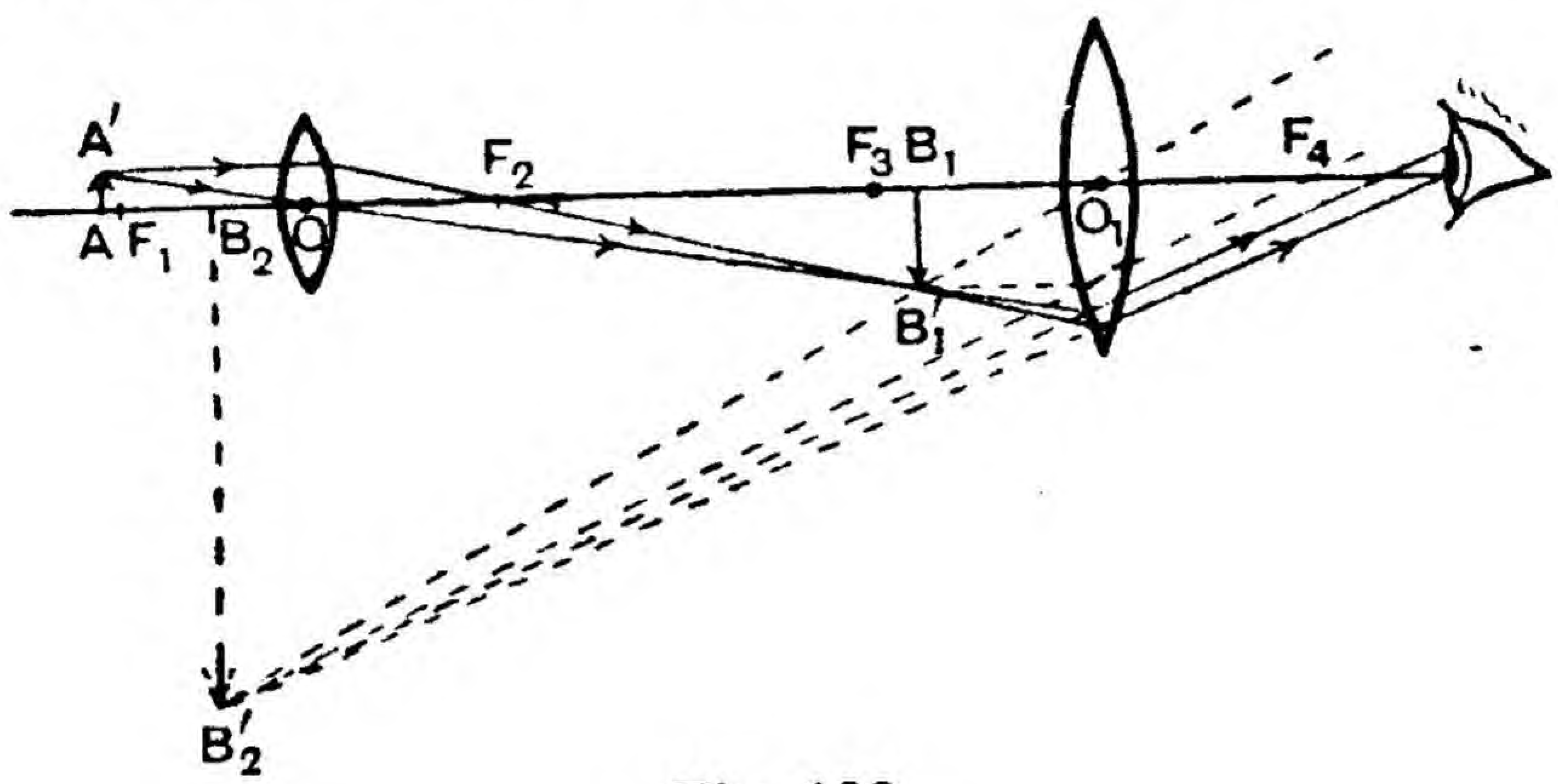


Fig. 132.

this distance when looking at a small object at a finite distance. The position of B_2' is fixed by drawing from B_1' the usual two construction rays, which are shown in dotted lines.

The magnifying power of the instrument can be calculated in the following way. If the length of the object is l, its greatest size as seen by the unaided eye is $\frac{l}{D}$, where D is the distance of distinct vision. If a_1 and b_1 are the distances of the object and image respectively from the objective, the length of the image B_1B_1' is $\frac{l\,b_1}{a_1}$. If a_2 and b_2 are the distances of this image and the final image respectively from the eyepiece, the length of the final image is $\frac{l\,b_1}{a_1}\cdot\frac{b_2}{a_2}$. If the eye is close up to the eyepiece, the angle which this image subtends at the eye is $\frac{l\,b_1}{a_1a_2}$. Hence the magnifying power of the microscope is

$$\frac{l\,b_1}{a_1a_2}\Big/\frac{l}{D}=\frac{b_1}{a_1}\cdot\frac{D}{a_2}$$

If the final image is formed at the distance of distinct vision, the above expression is also equal to the product of the *magnifications* of the objective and eyepiece. This is what we should expect, since, if the final image and object are situated at the same distance from the eye, the magnifying power

of the instrument reduces to the ratio of the length of the final image to that of the object, which is just the product of the *magnifications* of the two lenses.

Three important features of the microscope can be deduced from this expression for its magnifying power. Firstly, we can see that a very short focus objective is essential to a large magnifying power, for a good deal of the magnifying must be done by the objective, and so the ratio of $\frac{b_1}{a_1}$ must be, say, 50 if a magnifying of 500 is to be obtained. Thus the focal length of the objective must be about $\frac{1}{4}$ in. in order to keep the microscope from becoming unduly long. If it was made 1 in., for example, the length of the microscope would be over 50 in., which would make it quite impossible to manipulate the objects which are to be viewed. Secondly, we can see that it is essential that as much light as possible from each point of the object must enter the objective and pass through the microscope, otherwise the greatly magnified image will be too faint to be seen distinctly. This means that the diameter of the objective must subtend as large an angle as possible at the object, which involves making the objective with the greatest possible relative aperture. The smaller the diameter of a lens, the greater its relative aperture can be made without introducing too much aberration into the refracted pencils. So the objective is made with a very small focal length, since this allows of the greatest possible relative aperture and so the greatest brightness of the final image. Lastly, the above method of deriving the magnifying power of a microscope shows that there is no essential difference between it and the telescope. In each instrument the objective forms a real, inverted image of the object, whose size is greater than that of the object seen by the unaided eye, while the eyepiece magnifies this image after the manner of a magnifying glass. The focal length of the microscope objective is small merely to keep the instrument short and to increase the brightness of the image, but in essential principles the two instruments are the same. So it is not surprising to find that there are long focus microscopes and telescopes whose objectives are of comparatively small focal length, and that these two instruments are identical.

72. IMPROVEMENTS OF THE THIN LENS INSTRUMENTS

So far we have treated of optical instruments consisting of combinations of thin lenses, and the first telescopes were instruments of this sort, actual lenses which approximate very closely to the thin lens being used in their construction. We must now consider what defects will arise if the magnifying power of the instruments is increased, and how they may be eliminated or reduced to tolerable amounts. We will consider the telescope in the first place. We have seen (Art. 65) that the focal length of the eyepiece is fixed irrespective of the magnifying power, when the largest relative aperture permissible in the objective has been decided on. So increase in magnifying power can only be produced by

increasing the focal length of the objective and the problem arises of making objectives of focal length of 60 ft. and diameter 3 ft. for astronomical purposes. Since the field of view of an astronomical telescope is $\frac{M}{d_o+d_e}$, an increase of magnifying power will reduce the field of view unless the diameter of the eyepiece is increased in the same ratio as the focal length of the objective. This is quite impossible without having an eyepiece subject to very big aberrations of the oblique pencils, and so the field of view of a telescope decreases as the magnifying power increases, and an instrument of magnifying power of 50 will only have a field of about $\frac{1}{2}°$ on each side of the axis. In order to obtain needle-sharp definition over the whole of this field, the thin lenses must be replaced by combinations. The reader should contrast the problem presented by the telescope with that of the camera. Owing to the finite size of the "grains" of the photographic plate, the definition need not be as sharp as with the telescope, whose image is to be seen with the eye, which requires a smaller circle of least confusion for the best definition than the photographic plate. But the camera must cover a much wider field than the telescope to be a commercial product. So the camera lens gives fair definition over a wide field of some 50°, while the telescope must give needle-sharp definition over a small field of some 1° or 2°. This criterion limits the aperture of the objective to f/15 or, at the most, to f/10, but the small field means that the only aberrations which have to be eliminated are chromatic aberration, spherical aberration, and coma. This was originally done by using long telescopes in which the above aberrations are reduced to a minimum, but better methods are now available. Chromatic aberration is corrected out for the C and F lines in the case of telescopes to be used for visual work, since it is found by experience that the correction for these two colours gives the best results. This is done in the way described in Art. 43 by a suitable combination of a convex crown glass lens and a weaker diverging flint glass lens in contact. If the telescope is to be used in photographic work, the correction is made for three colours owing to the sensitivity of the plate to violet light, and so a triplet is necessary. Sufficient correction for spherical aberration is made by arranging for the lens combination to be convex-plane with the convex face pointing towards the object, so as to divide the total deviation produced by the objective equally between the two faces. The same disposition of the faces fortunately produces minimum coma. A reference to equations (32) and (33) in Art. 43 shows that $s' = \infty$, and so only *two* unknowns, r and s, are left to be uniquely determined by those equations.

73. EYEPIECES

It is evident that needle-sharp definition over the whole field will not be obtained with an eyepiece consisting of a thin lens, and we now turn

to the corrections to be made to this part of the telescope. It is well to realise that our task is made somewhat easier here by the fact that the pencil of rays from any one point on the object is a thin pencil by the time it reaches the eyepiece, and the reader will recall from the results of Chapter 6 that the diameter of the circle of least confusion due to the various aberrations is less for thin than for wide pencils. We have seen that the field of view of a telescope is bound to decrease as its magnifying power increases, if needle-sharp definition is to be maintained. As the biggest field of view is desirable, any way of increasing it is valuable, and a method has been found by using a compound eyepiece consisting of two converging lenses, one called the **field lens**, because it increases the field of view, and the other called the **eye lens**, because the eye is placed to it. In its simplest form the field lens is placed at the point where the real inverted image produced by the objective is situated. Fig. 133 shows three rays from a point on a distant object, which would

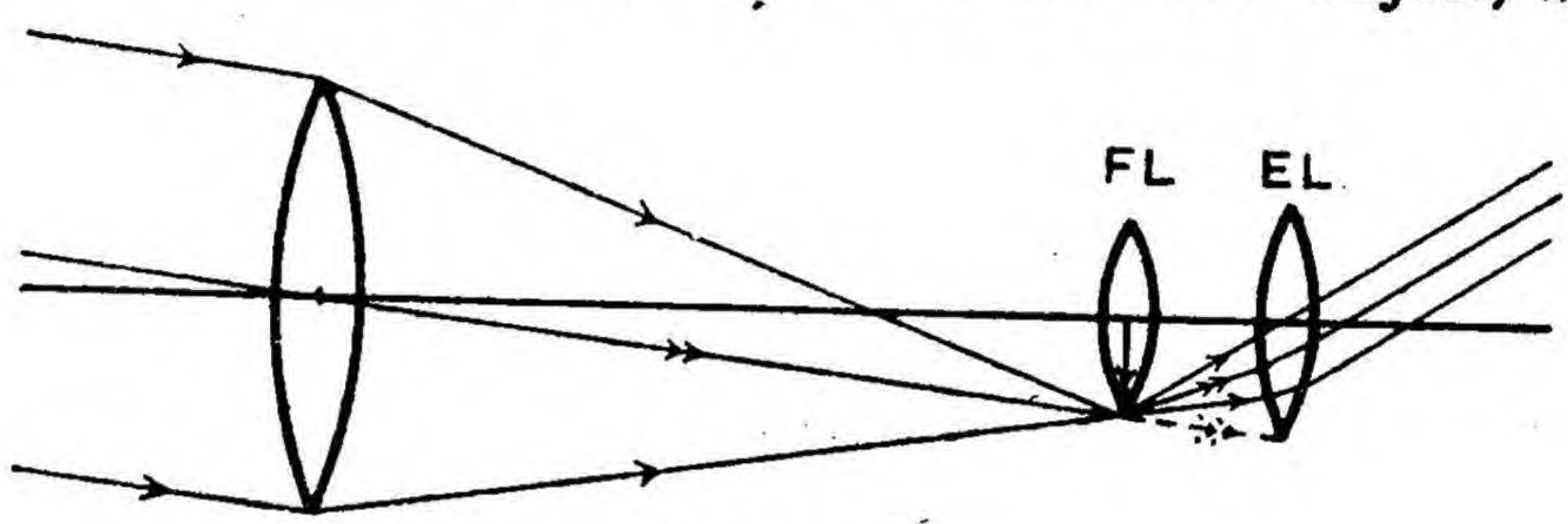


Fig. 133.

be on the edge of the field if the field lens were not employed. This follows from the fact that the central ray of the pencil, which is marked with a double arrow, would strike the edge of the eye lens in the absence of the field lens, this part of its path being shown in dotted lines. But the field lens bends the pencil inwards without affecting it in any other way, and so this point on the object will now be well inside the edge of the field. This insertion of the field lens does not affect the magnifying power of the instrument and so the field of view has been increased without introducing any drawbacks. This eyepiece has been modified and improved in two ways, leading to the Ramsden and Huygens eyepieces, which will be described briefly.

74. THE RAMSDEN EYEPIECE

The disadvantage of a compound eyepiece, in which the field lens coincides with the real inverted image formed by the objective, is that the eye lens is focussed on the field lens and so any scratches, or defects, on its surface are seen and spoil the look of the final image. This trouble is overcome in the Ramsden eyepiece by moving the two lenses closer together, so that the real, inverted image formed by the objective is produced before the field lens. The eyepiece consists of two plano-convex lenses of the same focal length with their plane faces outwards, the

lenses being separated by a distance equal to two-thirds of the focal length of either lens (Fig. 134). The real inverted image B_1B_1' formed by the objective is situated at the first focal plane of the eyepiece and the rays from each point of it are converged a certain amount by each lens, finally emerging as a parallel beam from the eye lens. The various aberrations are reduced to tolerable amounts in the following way: spherical aberration is reduced by sharing the deviations as equally as possible among four refractions instead of two; this helps to reduce coma as well. The radius of the circle of least confusion due to spherical aberration and astigmatism is also reduced by the fact that

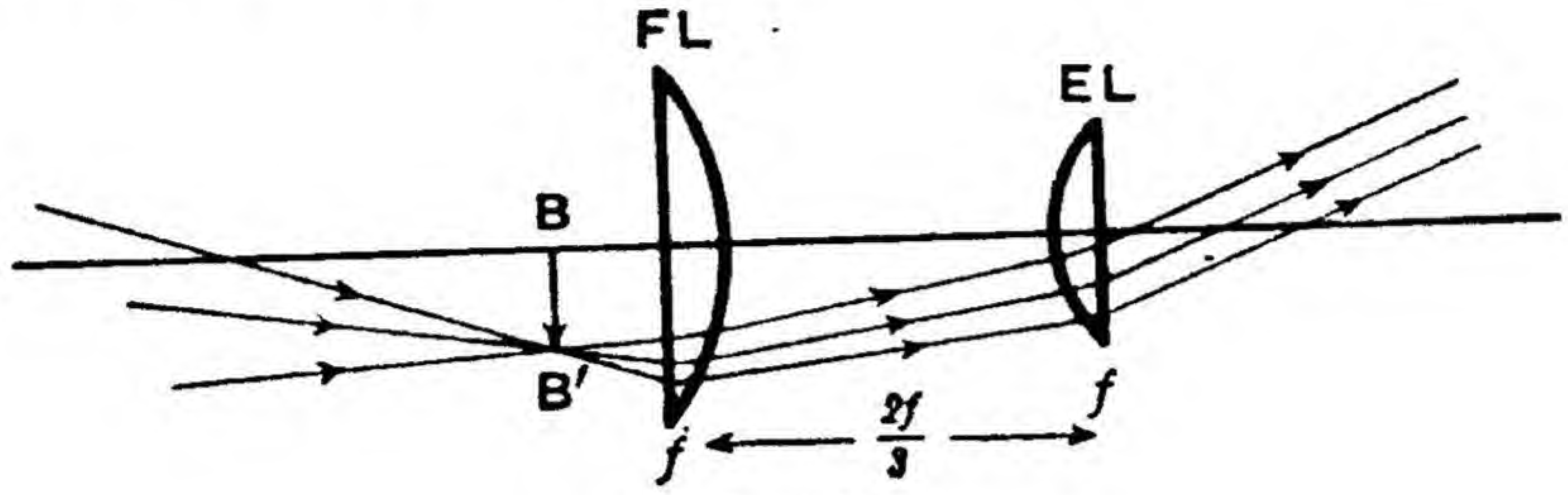


Fig. 134.

only narrow pencils pass through the eyepiece. Curvature of the field and distortion are also small over fields of 40°. Finally, chromatic aberration has to be eliminated in the sense that the size of the different coloured images is the same. This is the case for a combination of two lenses separated by a distance equal to half the sum of their focal lengths (Art. 43). This requires that the two lenses should be separated by a distance equal to the focal length of either of them, but as this would introduce the trouble mentioned at the beginning of the paragraph, the lenses are separated by only two-thirds of the focal length, thus sacrificing some elimination of chromatic aberration in order to avoid seeing any scratches on the field lens. The practical optician is continually faced with problems of this sort, in which the simultaneous elimination of two defects is impossible. He compromises by reducing one more than the other and the exact amount by which each is to be decreased to give the best results can only be found by practical experience. When this has been found, its realisation by the designer has to be attained by means of ray tracing.

75. THE HUYGENS EYEPIECE

The Huygens eyepiece was introduced in order to improve the corrections for both spherical and chromatic aberration of the Ramsden eyepiece. It consists of two plano-convex lenses with their plane faces turned away from the object (Fig. 135). The field lens has three times the focal length of the eye lens and the lenses are twice as far apart as the focal length of the eye lens, and so the correction for chromatic aberration is as good as possible. The real inverted image B_1B_1' formed by the

objective acts as a virtual object for the lens field which forms a real image of it at B_2B_2' at the first focal plane of the eye lens. So this lens forms a final image of the object at infinity behind the telescope. Since this eyepiece requires a virtual object to be presented to it, it cannot be used with cross-wires as the Ramsden eyepiece can. It is often called a negative eyepiece to distinguish it from the Ramsden eyepiece, which is called a

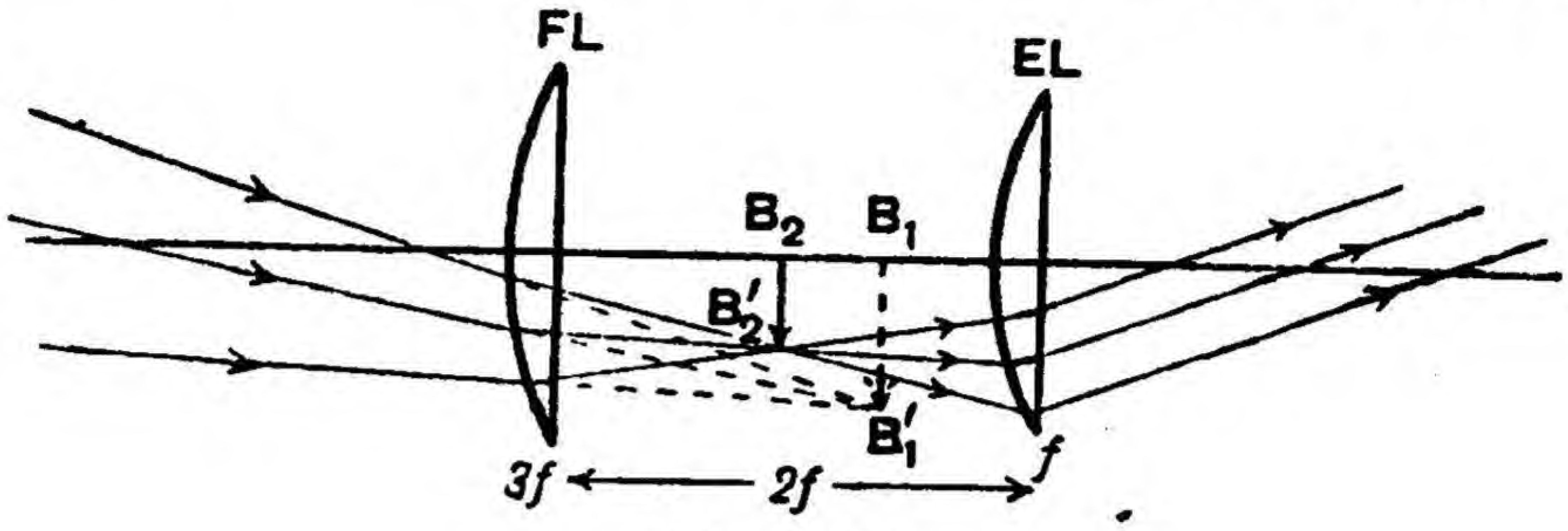


Fig. 135.

positive eyepiece. It has already been mentioned that the correction for chromatic aberration is as good as can be achieved, and the correction for spherical aberration and coma is better in this eyepiece than in the Ramsden eyepiece. It is also free from astigmatism and distortion to the same extent, but it suffers from considerable curvature of the field, the curvature being convex towards the observer's eye.

76. MICROSCOPE OBJECTIVES

In a microscope which magnifies 500 times, the objective alone produces a magnification of 50, so that it will need to be very carefully corrected for the various aberrations, if a reasonably sharp image is to be obtained. Spherical aberration, coma, and chromatic aberration must be reduced to a minimum, the remaining aberrations being unimportant since the field of view of a microscope is so small. At such high magnifications as 50, satisfactory correction can only be obtained for one image distance, and in the best objectives use is made of the aplanatic surface to obtain a very wide angle pencil from each point of the object. The reader will recall (Art. 46) that if a point object, A (Fig. 136), is placed at a distance $\frac{n_2}{n_1}r$ from the centre C of a spherical surface of radius r separating two media of refractive index n_1 and n_2 $(n_1 > n_2)$, a geometrical point image B is formed at a distance $\frac{n_1}{n_2}r$ from C, however great

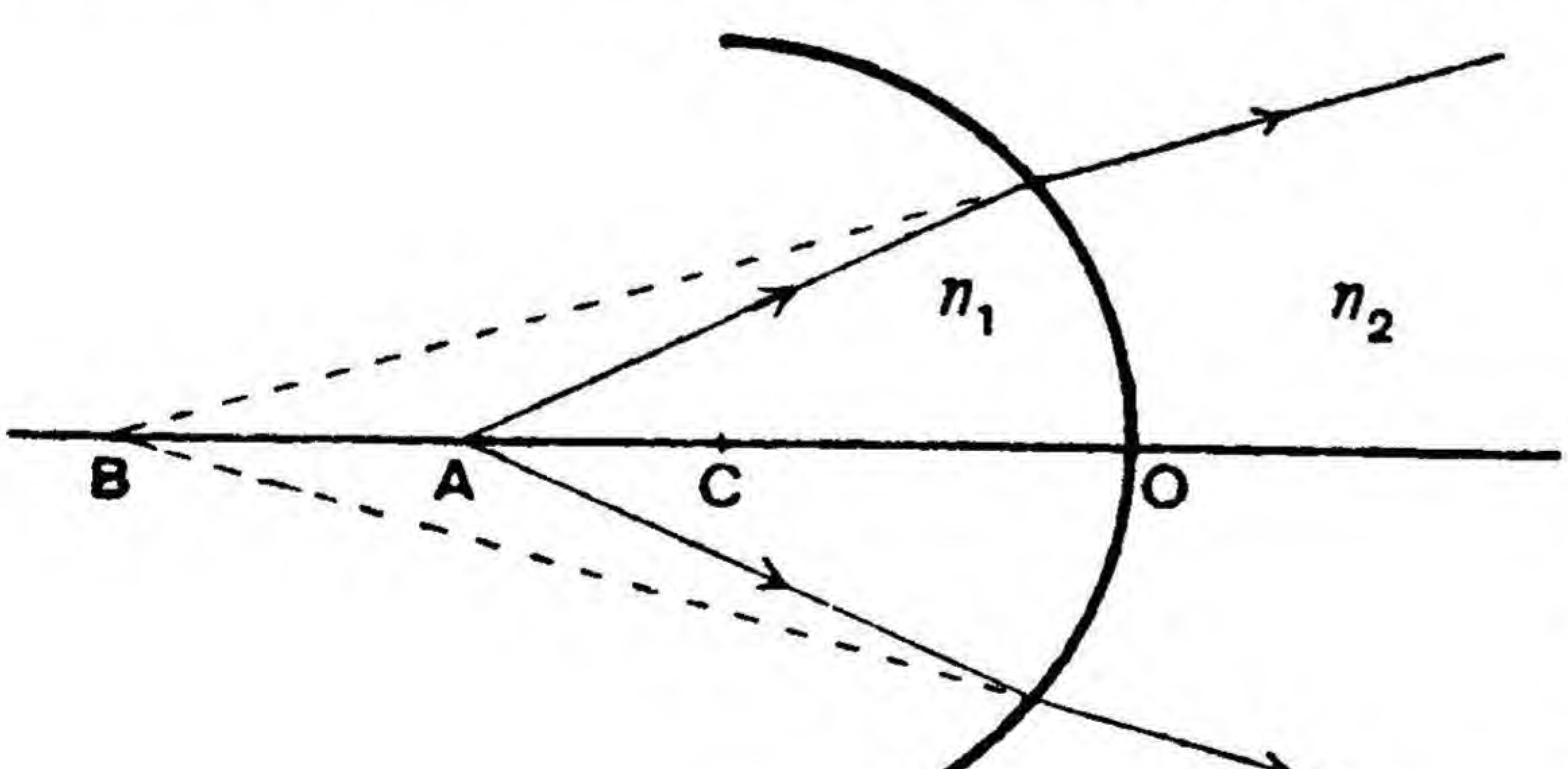


Fig. 136.

the angle which the rays from A may make with the axis of the refracting surface. In other words, the image is entirely free from both spherical aberration and coma. A and B are called the **aplanatic points** of the refracting surface. This special case is made use of in microscopic objectives in the first two stages of the refraction. The object A (Fig. 137) cannot be embedded in the glass of the first lens, so it is immersed in a film of oil below the first lens having the same refractive index as the lens. The first lens is hemispherical and the object is placed at one of the aplanatic points, and this lens forms a geometrical point image of the point object at the other aplanatic point B_1. This is arranged to be at the centre of curvature of the lower face of a meniscus lens L_2 so that there is no refraction at this surface. It is also the aplanatic point for the second

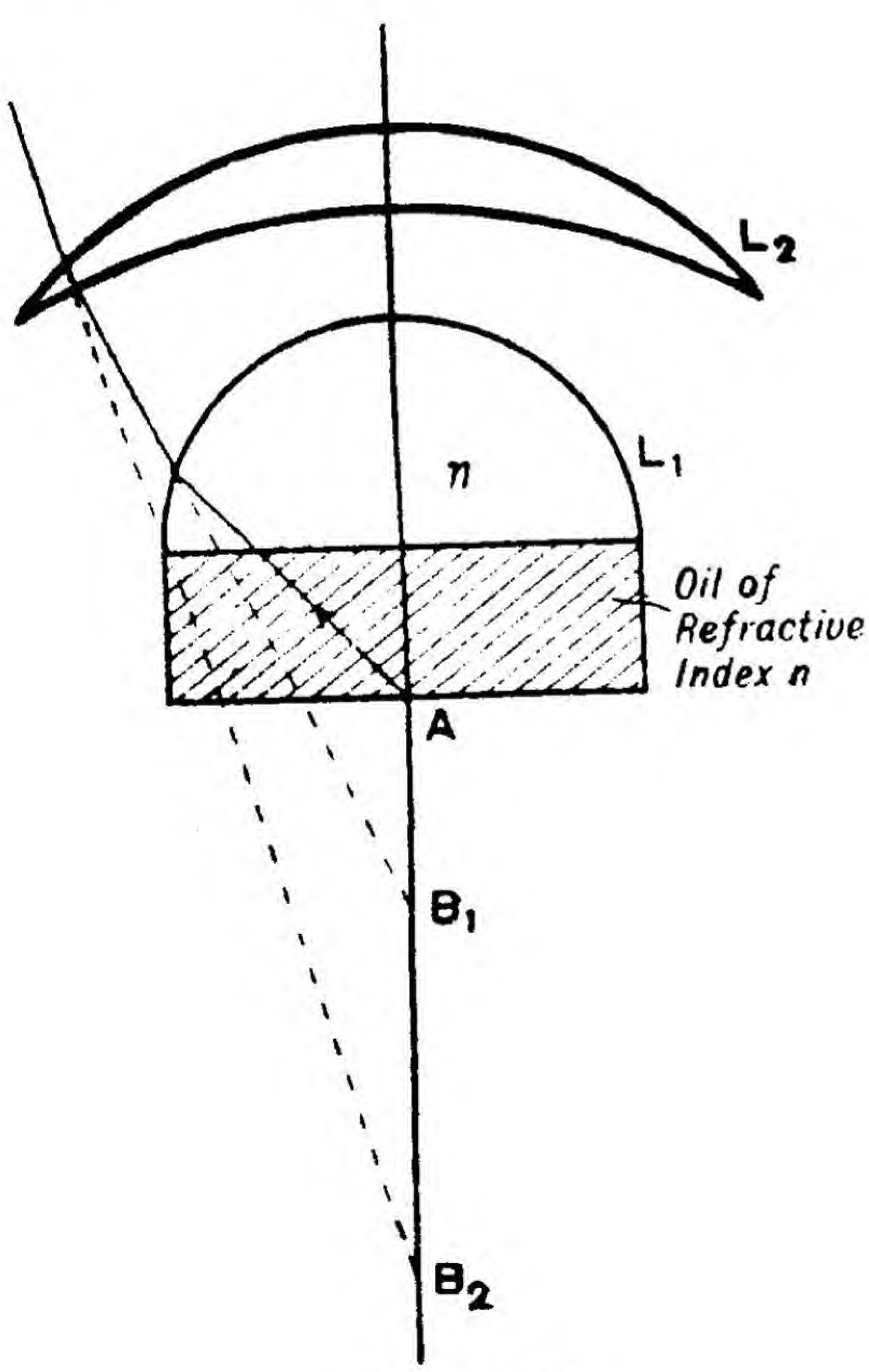


Fig. 137.

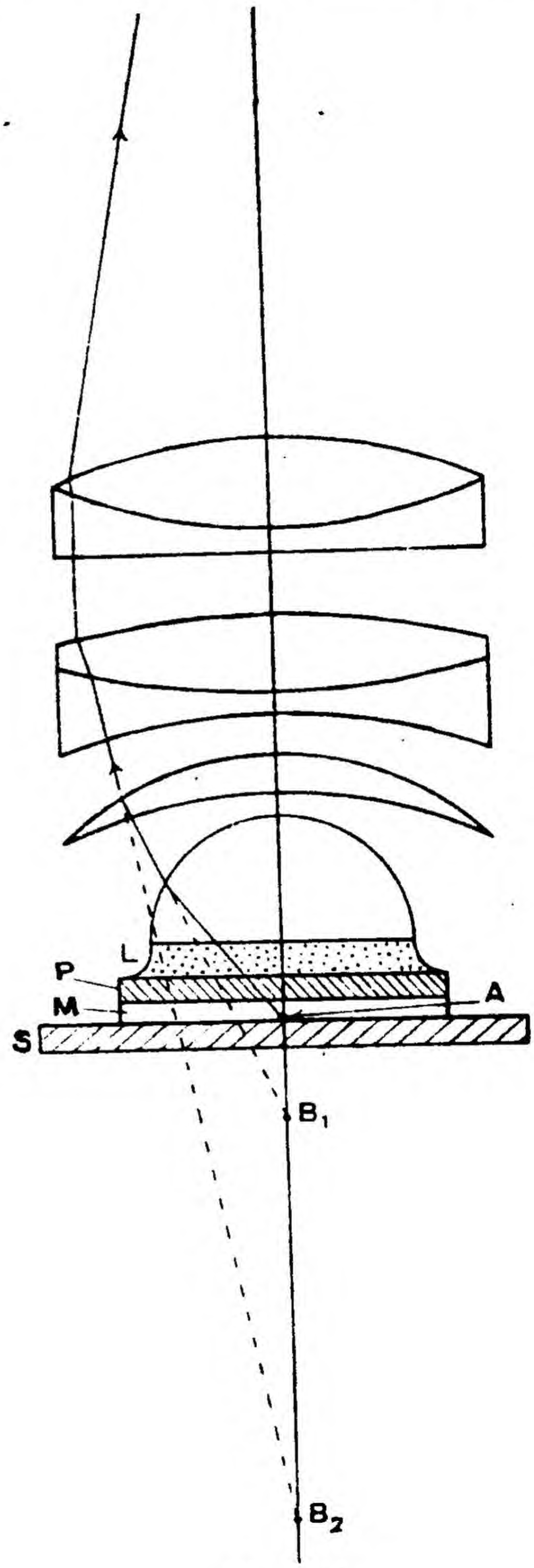


Fig. 138.

surface of the lens, and so this lens produces a perfect point image of B_1 at B_2. So the original wide angle pencil from the point object A has now been transformed into a pencil of smaller angle diverging

from the point image B_2, and this diverging beam is brought to a good enough focus by two lens combinations, which minimise spherical aberration and coma in the usual way. It is evident that the first two lenses of the complete objective, which is illustrated in Fig. 138, will introduce some chromatic aberration, and this must be corrected to some extent by the last two components of the system, which are not therefore simple achromatic lenses. This type of objective is known as an **immersion achromat** and corrects for spherical aberration at one colour and chromatic aberration at two colours. The **apochromatic objective** consists of ten lenses and corrects for spherical aberration at two colours and chromatic aberration at three; but an account of this objective would be out of place in this book, and those readers who are interested enough to pursue this subject further should consult Hardy and Perrin's "Principles of Optics," or Martin's "Introduction to Applied Optics," volume 2. The Huygens eyepiece is most commonly used in the microscope, although it is usual to employ the Ramsden eyepiece in travelling microscopes where cross-wires must be fitted.

77. THE PROJECTION LANTERN

The purpose of the projection lantern is to cast a magnified image of a lantern slide of some object on a screen. In the cinematograph projector a greatly magnified image of a representation of some object on a film is cast on to the screen, sixteen such images being thrown on the screen in every second. In the epidiascope one lens serves for the projection of lantern slides, while the other serves for the projection of an image of an opaque object, which is suitably illuminated.

We shall consider the projection of lantern slides in the first place. Suppose that the object to be projected is a set of black and white squares,

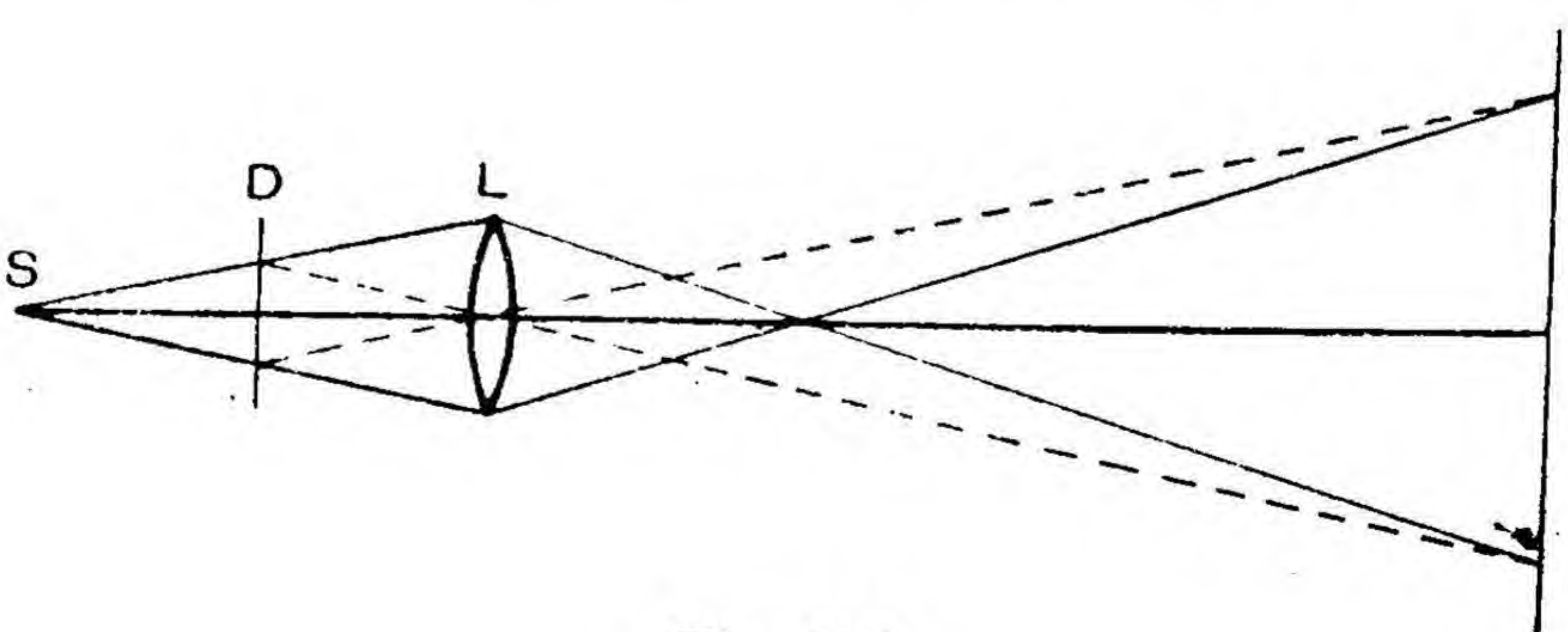


Fig. 139.

such as a chess-board. Then the lantern slide would be a small-scale representation of the board on a glass plate, the white squares being transparent to light and the black ones being opaque to light. In order to project an image of the slide D on to the screen R (Fig. 139), a beam of light from a source S must be sent through it and a suitable converging lens L must be placed between the slide and screen to form the magnified

image. The standard size of lantern slides in this country is $3\frac{1}{4} \times 3\frac{1}{4}$ in. and the diagram shows quite clearly that only about a quarter of this area will be projected on to the screen by a lens of diameter 3 in., which is considerably greater than that of the ordinary projection lens. In order to increase the field of view of the projection system, a converging lens N, called the **condenser**, is placed just in front of the slide on the same side as the source of light and its focal length is so arranged that it produces a real image of the source at the centre of the projection

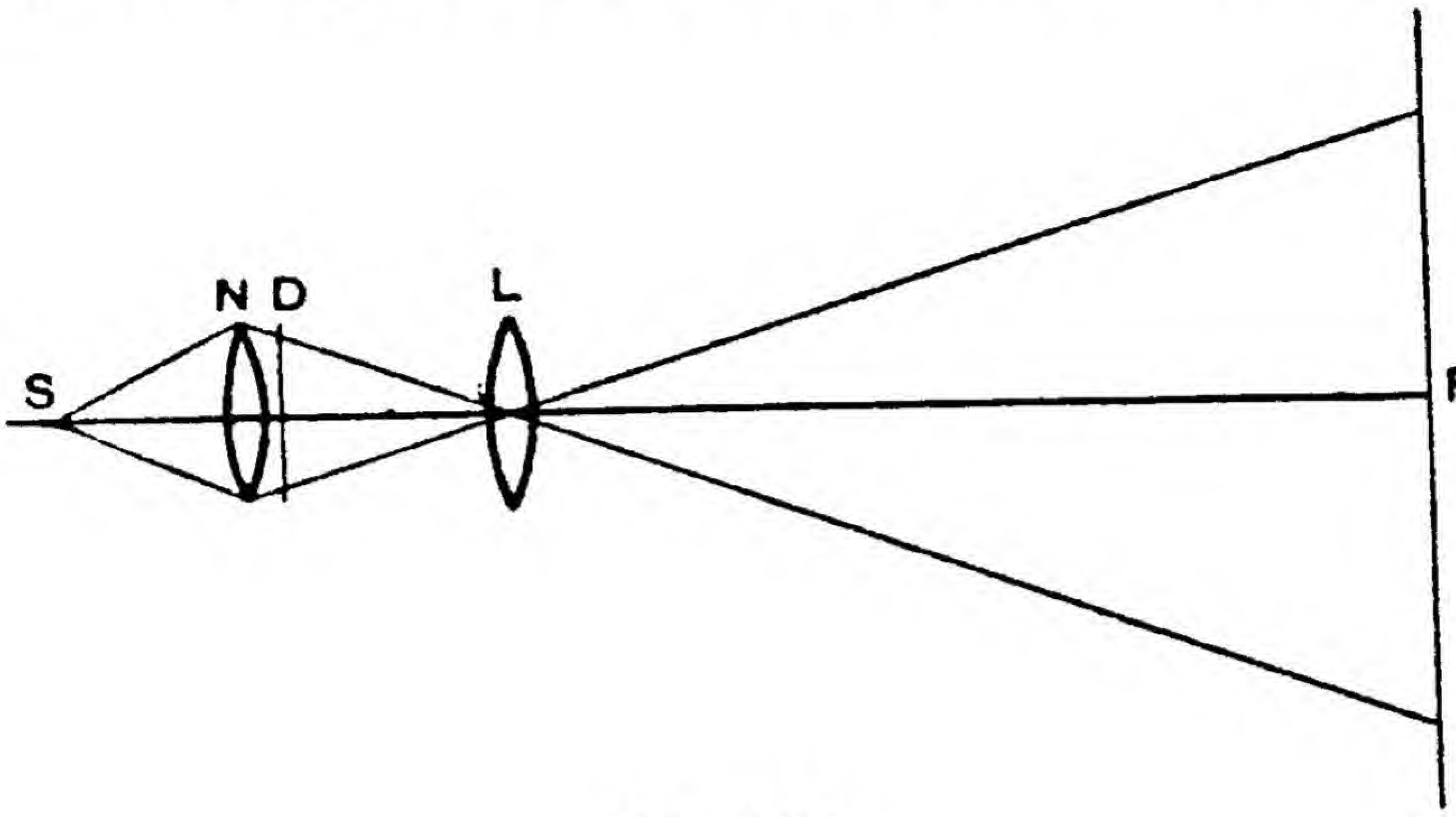


Fig. 140.

lens. It will be quite clear from Fig. 140 that this greatly increases the field of view of the projection system. An examination of Fig. 141 will also make it clear that the condenser must be free from spherical and chromatic aberration, otherwise the field of view is not quite so large as it could be in the absence of these aberrations. For the marginal rays passing through the condenser in Fig. 141 do not pass through the projection lens if it is in the position shown, and so the outer portions of the slide will not be seen on the screen. Therefore the condenser is

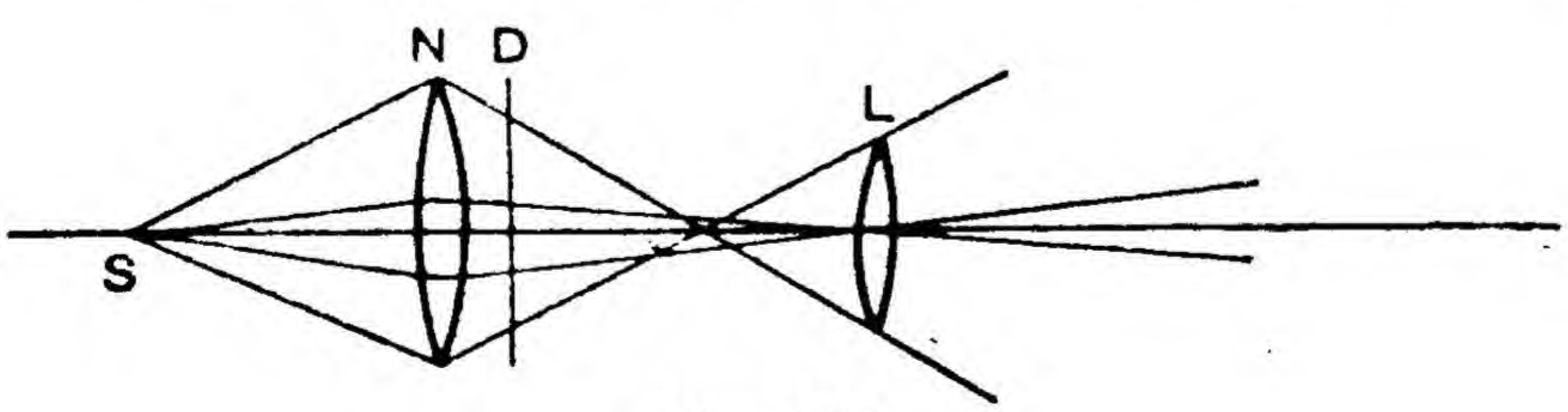


Fig. 141.

made of two plano-convex lenses with their plane faces turned outwards so as to reduce spherical aberration to a minimum by sharing the total deviation equally between the four refractions, and the complete projection system is shown in Fig. 142. The projection lens itself is the same as lenses used in portrait photography, since it is merely a portrait lens working backwards. The requirements of the two lenses are the same, namely, good definition over a moderate field of some 30° in all, which is sufficient to give a picture 6 ft. square on a screen 12 ft. from the lantern ; a lens of large aperture to give the maximum amount of light

on the screen; finally, no depth of focus. It is made of three or four components and more advanced books must be consulted for details of the various types. The chief difficulty in the episcopic projection of opaque objects is to get a really bright image on the screen, and this has to be done, in the first place, by producing the greatest possible illumination on the opaque object by using a concentrated filament high-powered lamp and reflecting as much light as possible from it on to the object, and secondly by using a projection lens of the greatest possible aperture so as to collect the greatest amount of light from each point of the object. Therefore the projection lens for episcopic projection has a greater diameter than that of the lens used for lantern-slide projection.

The development of optical instruments is an admirable illustration of how progress is most rapid in scientific achievement when theory and experiment go hand in hand, and how pure science not only revolutionises industry by the discovery of new ideas which lead to quite new commercial processes, but is in its turn stimulated to the discovery of new knowledge by the demands of industry. We have seen how the wonder-

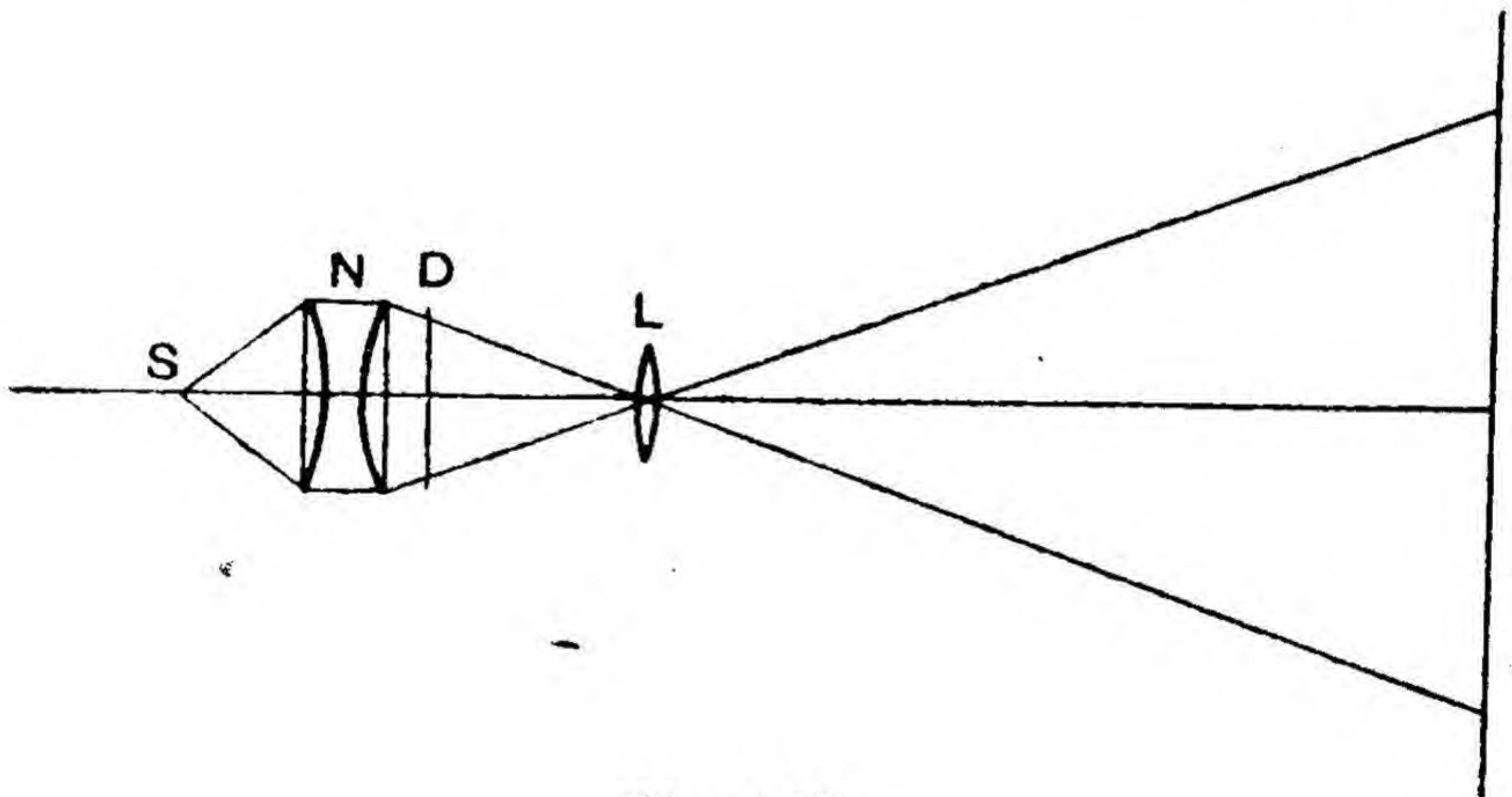


Fig. 142.

ful properties of the ancient natural lenses and mirrors led to the desire to understand their powers, and so to the foundation of geometrical optics and the conception of the thin lens and theorems concerning it and spherical mirrors. These theorems threw light on the working of the telescope and microscope and led to their improvement, but it was the very imperfection of the biggest telescopes of his time that led Newton to do his work on colour. He diagnosed the cause of the trouble correctly but was unable to find a remedy; it was in the attempt to follow up Newton's work that Fraunhofer discovered the dark lines in the solar spectrum and so laid the foundations of the new science of spectroscopy, which has revolutionised the very basis of theoretical physics. After chromatic aberrations had been greatly reduced, a demand for still further improvement in the performance of optical instruments called for further development of geometrical optics, which led to the discovery of the various aberrations of oblique pencils, which has minimised the labour of the lens designer by reducing the field over which

he must apply the method of trial and error used in ray tracing. It is this happy combination of theory and experiment, of logic and empiricism, of science and industry, which has led to the excellent performance of modern optical instruments.

EXAMPLES ON CHAPTER VIII

1. A newspaper is held 50 cm. from the eye and viewed through a convex lens of focal length 100 cm. Find an expression for the apparent magnification in terms of the distance of the lens from the newspaper, and deduce where the lens should be placed to give the maximum apparent magnification. (*Oxford Schol.*)

2. Explain the principle governing the design of a simple astronomical telescope and trace the path of a typical ray through the instrument.
If the moon is viewed through such an instrument and half of the object glass is covered, what change would you expect to find in the image? (*Oxford Schol*).

3. Describe some form of telescope and explain how you would determine its magnifying power experimentally.
A telescope is used to view distant pairs of objects, such as stars. Explain in general terms why it is impossible to recognize the presence of two objects if their angular separation is less than a certain amount. What factors determine this minimum? (*Camb. Schol.*)

4. Explain the action of a telescope made with two convex lenses, illustrating your answer by a diagram showing the paths through the system of a pencil of light from a point on the object which is not on the axis.
Explain what is meant by the magnifying power of the telescope, and show how it depends on the focal lengths of the lenses. (*O. and C.*)

5. Two convex lenses each of 20 cm. focal length are set up coaxially 5 cm. apart. An image of a flagstaff 200 m. distant and 10 m. high is formed by the combination. Find the position and size of the image. (*O. and C.*)

6. What is meant by the magnifying power of (a) a telescope, (b) a microscope?
Deduce a value for the magnifying power of a telescope if both the object and image are at infinity.
A telescope consisting of an objective of focal length 10 in. and an eyepiece of focal length 1 in. is used to view an object 100 ft. distant. If the final image is seen at infinity, find the magnifying power of the instrument. (*O. and C.*)

7. An astronomical refracting telescope is adjusted to give a real image of the sun upon a screen. Draw a diagram showing the path of a pencil of rays through the telescope to a point on the boundary of the image.
If the focal lengths of the object glass and eye lens are 100 cm. and 2·5 cm. respectively and the image of the sun, formed on a screen placed 30 cm. from the eye lens, is 9·6 cm. in diameter, find the angle which the sun subtends at the centre of the object glass. (*N.U.J.B.*)

8. Describe the optical arrangement of a simple astronomical telescope. Draw a diagram showing the paths through the telescope of two rays from a distant point not on the axis.
If the focal lengths of the objective and eye lens of such a telescope are 60 cm. and 5 cm. respectively, and the distance between the lenses is 64·5 cm., where will the final image of a distant point be formed? (*N.U.J.B.*)

9. Write an essay on the optics of the refracting telescope. (*N.U.J.B.*)

10. A spectator is sitting on a line bisecting a cricket pitch at right angles. The batsman 6 ft. high is 150 ft. from the objective of the astronomical telescope through which the spectator is looking; the focal length of the objective is 10 in., that of the eyepiece is 2 in. and its diameter is 0·4 in. If the telescope is adjusted to produce the final image at infinity, find the angle subtended by the image at the

eye, the angle subtended by the batsman at the objective, and the magnifying power of the telescope. What length of the pitch can be seen in the telescope? What adjustment must be made to bring the final image to a distance of 10 in. from the eye and what will be the magnifying power of the telescope now?

11. A Galilean telescope is made from a concave lens of 2 cm. focal length and convex lens of 20 cm. focal length. Make a drawing to scale of the paths of typical rays when the instrument is directed towards a distant object.

(*Oxford Schol.*)

12. Describe and explain, with diagrams, the optical system of Galileo's telescope.

Explain why, in this telescope, the concave lens produces a magnified image.

(*O. and C.*)

13. Describe with the aid of a diagram the construction and mode of action of a Galilean telescope. What advantages has this system over the astronomical telescope?

The objective and eyepiece of the Galilean telescope have focal lengths of 25 cm. and 5 cm. respectively. Calculate the angular magnification produced when the system is used to view an object at a distance of 4 metres from the objective, the final image being formed at the minimum distance of distinct vision (25 cm.) from the eyepiece.

(*O. and C.*)

14. A pair of opera glasses has objectives of 4 in. focal length and eyepieces of -2 in. focal length. They are used to look at a pianist 100 ft. away in a concert hall, the final image being produced at 12 in. from the eye. If the pianist's hand is 8 in. long, what will be the length of the image and the magnifying power of the opera glasses? If he stands up, what length of him can be seen in the glasses, if the diameter of the objective is 0·7 in.?

15. Describe the reflecting telescope. What are the advantages of this type of telescope? A concave mirror 30 ft. focal length is used to form an image of the moon. What will be the diameter of the image if the moon subtends an angle of $\frac{1}{2}°$ at the eye? How does the brightness of the image depend on the diameter of the mirror?

(*Oxford Schol.*)

16. Describe the optical system of a reflecting telescope, giving a diagram showing the paths of three selected rays from the same point not on the axis. Is the image upright or inverted in the case you describe? (*O. and C.*)

17. A telescopic objective of focal length 5 ft. is to be achromatic for the C and F lines, and spherical aberration is to be reduced to a minimum by making the surface facing the eyepiece plane, and the surface facing the object convex. The crown and flint glass components are to be cemented together. Find their focal lengths and the radii of curvature of their surfaces, using the table of refractive indices given in Chapter V.

18. An astronomical telescope of magnifying power 300 is to be designed. The aperture of the objective must not exceed 1/12 of its focal length and that of the eyepiece $\frac{1}{3}$ of its focal length in order to keep aberrations to a minimum. Calculate the focal lengths of the objective and eyepiece, if the diameter of the pupil of the eye is 0·1 in. Find also the field of view of the telescope.

This telescope is used to look at the moon, which subtends an angle of 1° at the earth, the final image being at infinity. Find how much of the magnifying power of 300 is due to the objective and how much to the eyepiece.

19. Draw a diagram showing the path of rays through a compound microscope from an object point not on the axis, when the image is formed coincident with the object at the distance of most distinct vision from the eye.

If this distance is 28 cm., the distance of the object from the objective 4 cm., and the magnifying power of the instrument 14, find the focal length of the lenses.

(*Camb. Schol.*)

20. Describe the compound microscope, illustrating your answer by a diagram showing the passage of rays through the system. Distinguish between the magnification and the resolving power of a microscope and state on what factors each depends.

(*Camb. Schol.*)

21. Describe the optical system of a microscope, and draw the path of the rays concerned when it is used to view a linear object. What is an ultra-microscope? *(London.)*

22. Two converging lenses of focal lengths 4 cm. and 5 cm. respectively are placed 20 cm. apart and used with the 4 cm. lens as the objective and the 5 cm. lens as the eyepiece of a microscope. Where must the object be placed, so that the final image may be 25 cm. from the eye and virtual? *(Camb. Schol.)*

23. Discuss the difference between a modern refracting telescope and the "thin lens" telescope. State clearly what aberrations must be minimised in each part of the instrument and how this is achieved in each case. Explain also how the field of view has been increased.

24. Discuss the properties of the ideal microscopic objective and explain how and to what extent they are realised in practice.

25. Optical instruments always reduce the brightness of finite luminous objects. Discuss this statement. *(Oxford Schol.)*

Chapter IX

PHOTOMETRY

78. INTRODUCTORY

Since the industrial revolution took place, there has been a steady growth in the population of the great nations of the world and an equally steady migration of population from the country to the towns, resulting in the production of a considerable number of very large towns with populations of hundreds of thousands. The industrial revolution has also been responsible for the birth of many new industries, such as cotton spinning, the wool industries, the iron and steel industries, which employ thousands of people in indoor occupations. Consequently there has recently arisen a new interest in problems of artificial lighting. What is the correct amount of light in factories in which people work? What is the correct illumination for the ledgers in which clerks write and on the blue prints at which mechanical draughtsmen labour? What is a suitable amount of illumination in the streets at night, so as to render them safe for the pedestrians who walk along them and the motorists who drive through them? And when the correct illumination has been decided upon, how is it to be produced? Will the lamps at present at the disposal of the illumination engineer be satisfactory, or is it necessary to produce new and more efficient types? And how are the lamps to be arranged so as to give the required illumination of the surface without any shadows and without glare? These are some of the main problems which confront the illumination engineer in this age of large towns, huge factories, and rapid motor traffic. It is evident that they can only be solved by the scientific method of learning to *measure the various quantities in numbers*; once this has been done, it is possible to make some progress, and we shall see that considerable progress has been made. It is evident that two quantities have to be measured in numbers. The first refers to sources of light. We should all agree that an ordinary electric light, such as is used in a living-room, is more powerful than a small pea lamp in an ordinary flash lamp. By this we mean that the electric lamp appears more powerful than the pea lamp when they are the same distance away from our eye. Presumably the electric lamp sends more light per second into our eye than the pea lamp. But we want to be able to measure both amount of light and this power of the lamp in numbers in some way. The scientific name for the amount of

light entering the eye or passing any plane in one second is the **light flux**, and the scientific name for the power of a source of light is its **luminous intensity**. The second quantity refers to the brightness of a surface which is being illuminated either by daylight or by artificial light. For example, a person reading a book as twilight is coming on will move towards the window, since the page of his book will be more brightly illuminated. The scientific name for the brightness of a surface is the **illumination** of the surface, and we shall now discuss how these three quantities and any others arising out of them may be defined and measured. After that we shall conclude the chapter with a brief account of the solution of the problems we have mentioned above.

79. FUNDAMENTAL QUANTITIES

It is well known that light is a form of energy, for, if the rays of the sun are concentrated by a lens on to a piece of paper, they will set it on fire, proving that light can be converted into heat, a form of energy, and so that light is itself energy. So it is natural to define the light flux through a given plane as the number of ergs of energy in the form of light which cross the plane in one second. This would be the most logical definition, but it has not been adopted for two reasons. In the first place a given number of ergs of light entering the eye per second produces different degrees of sensation according to the colour of the light, yellow producing the greatest effect and other colours proportionately less, as the limits of the visible spectrum are approached. Consequently the number of ergs per second of light flux does not give any real idea of the magnitude of the sensation produced in the eye until the colour is specified. Secondly it would be inconvenient to have such a unit for practical purposes, because it is rather laborious to measure the light flux in ergs per second, and it is further unsatisfactory because its value is so small. For example, the eye can just detect a light flux of $1{\cdot}7\times10^{-9}$ ergs per second; an ordinary electric lamp consuming electrical energy at the rate of 60 watts gives out light at the rate of 1·67 watts. So the unit of light flux has been defined in quite a different way. In the first place a standard source of light, the **standard candle** of constant luminous intensity, has been agreed upon and the unit of light flux has been derived from it. This is far more convenient in practice and bears a much closer relation to the sensation produced in the eye by the light.

A satisfactory standard source of light must give out a constant amount of light per second, it must be reasonably easy to make and be accurately reproducible. This last statement implies that two competent physicists in different parts of the world must be able to make the source and that, if they bring their sources to the same place for comparison, they will have the same luminous intensity to the required degree of accuracy, which is 1 in 1000 in photometry. The first standard suggested was the

candle, which was made of a specified kind of wax, of a definite diameter and mass, and was to burn at a definite rate. It was abandoned because its luminous intensity was found to vary with the length of the wick and other factors, and so it failed in the matter of constancy. It was replaced by the Vernon Harcourt pentane lamp, which has a luminous intensity of 10 times the candle when it burns in air at 760 mm. pressure containing 8 parts in 1000 of water vapour. This standard is now accepted internationally. It is somewhat difficult to set up and so attempts have been made to find more suitable standards. In Germany the Hefner amyl acetate lamp is used and has a luminous intensity of 0·90 candles when burned in the standard way, and it has been suggested that the luminous intensity of a square centimetre of platinum at the temperature of its melting-point should be adopted ; but it has not yet been possible to realise this to the required standard of accuracy and reproducibility. But, to avoid the length of time needed to set up and operate a Vernon Harcourt pentane lamp, both the National Physical Laboratory in this country and the Bureau of Standards in America have adopted electric lamps of standard construction and burning at a definite wattage as sub-standards, as they are so much quicker to set up and easier to use. The sub-standards are compared periodically with the pentane lamp. So the unit of luminous intensity is the **international candle, which is a point source emitting light uniformly in all directions at one-tenth of the rate of the Vernon Harcourt pentane lamp burning in air at 760 mm. pressure containing 8 parts in a 1000 of water vapour.** It is evident that this source is a theoretical concept only.

The unit of light flux is the **lumen, which is the amount of light falling in one second on a unit area placed at right angles to the direction of the light at unit distance from an international candle.** The reader must be careful to realise that the lumen measures the rate at which light energy passes through or falls on the area ; it does not measure a quantity of light, but a rate of flow of light. And we may add here that the term, amount of light, is not a vague one ; it means the quantity of light energy and the dimensions of the lumen are those of power.

The **luminous intensity of a source is the amount of light falling in one second on unit area placed at unit distance from the source at right angles to the direction of the light.** For example, if the luminous intensity of a source is 50 candle-power, it means that the source produces a light flux of 50 lumens on the unit area at unit distance, which is 50 times that produced by an international candle. This is the reason for calling the unit of luminous intensity the **candle-power.**

The illumination of a surface is the amount of light falling in one second on unit area of the surface.. The unit of illumination is either a lumen per square foot or the **foot-candle.** This is the illumination on a surface 1 ft. from an international candle and is obviously equal to a lumen per square foot. Other units of illumination are the

metre-candle, the illumination on a surface at a metre from an international candle, and the phot, the illumination on a surface at 1 centimetre from an international candle. In each of these three cases the surface is at right angles to the direction of the light falling on it from the source. It is evident that a metre-candle is equal to a lumen per square metre, the phot to a lumen per square centimetre, so that there are 10,000 metre-candles in a phot.

Finally the **brightness** of an extended source of light is defined as the candle-power of unit area of the source. It is important to realise that the term, luminous intensity, is restricted to point sources of light.

80. FUNDAMENTAL LAWS : THE INVERSE SQUARE LAW

Before we can go on to the various ways of measuring luminous intensity and illumination, we must establish some fundamental laws relating these two quantities. It is possible to do this theoretically provided that we assume that our sources of light are point sources radiating light equally in all directions. It is therefore necessary to bear in mind that actual sources may not fulfil these conditions exactly, and we must be prepared to find evidence of this in our experimental results and to make the necessary allowance for it. Let us imagine, then, that we have a point source of candle-power P radiating light equally in all directions at the centre of a sphere of radius r (Fig. 143). The total light flux emitted by this source is 4π P, since it sends P lumens on to a unit area at unit distance and the area of a sphere of unit radius is 4π. Hence the total light flux falling on the sphere of radius r is 4π P, and so the illumination I of the sphere is given by

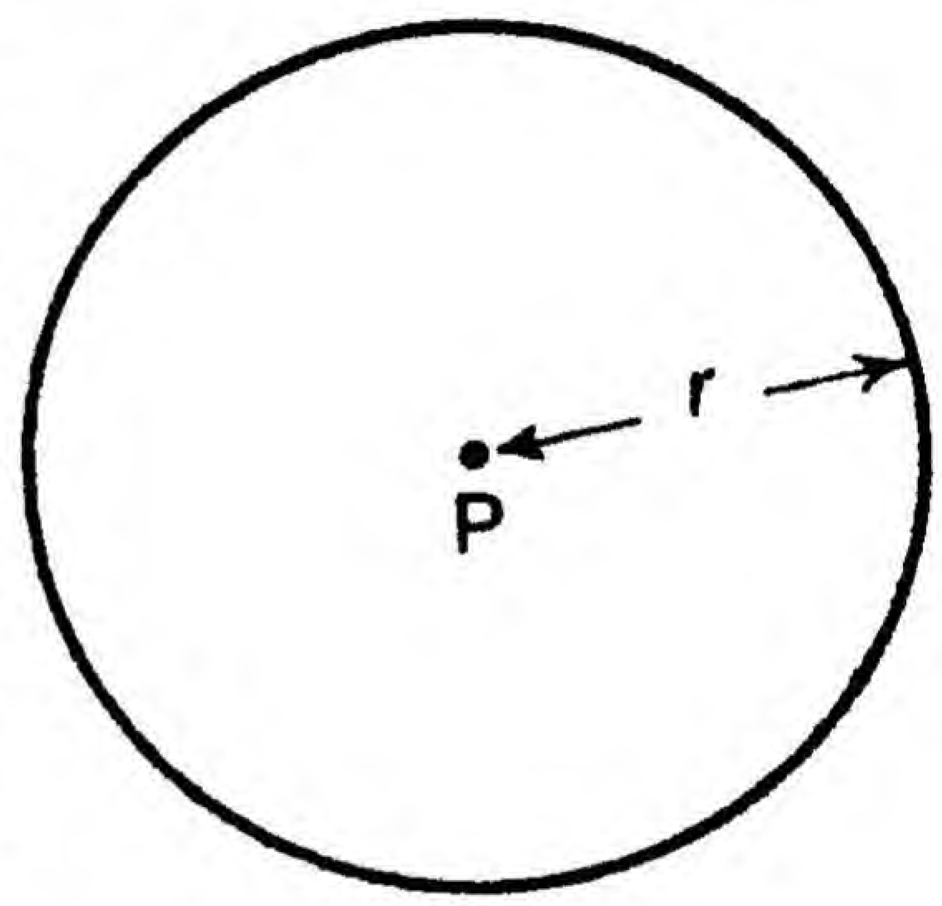


Fig. 143.

$$I=\frac{4\pi P}{4\pi r^2}$$

$$\therefore I=\frac{P}{r^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

Hence we see that **the illumination of a surface is inversely proportional to the square of its distance from a point source of illumination**. This is known as the **inverse square law** and is due to Lambert.

It is also important to realise that the illumination of a surface due to a given light flux changes as the angle between the direction of the light flux and the normal to the surface changes. Suppose there is a light flux L lumens falling normally on the surface AB (Fig. 144), whose length is unity in a direction normal to the plane of the paper. Then the illumination I of the surface is given by

$$I=\frac{L}{AB}$$

Now consider the illumination of the surface BC, whose normal makes an angle θ with the direction of the light incident on it. The light flux falling on the surface is the same as before, L lumens, but the area it covers is now BC, so that the illumination of the surface BC is given by

$$I_\theta = \frac{L}{BC} = \frac{L}{\frac{AB}{\cos\theta}} = \frac{L}{AB}\cos\theta$$

$$\therefore\ I_\theta = I\cos\theta \qquad (47)$$

Hence the illumination of a surface varies as the cosine of the angle between the direction of the light flux illuminating it and the normal to the surface, and so decreases as the surface is turned to be more oblique to the incident light. It is important to realise here that the illumination of a surface is not the same thing as its brightness ; the one refers to the amount of light *falling on* unit area of it in unit time and the other to the amount of light *emitted by* unit area of it in unit time. So the brightness will only be equal to the illumination provided that the surface reflects

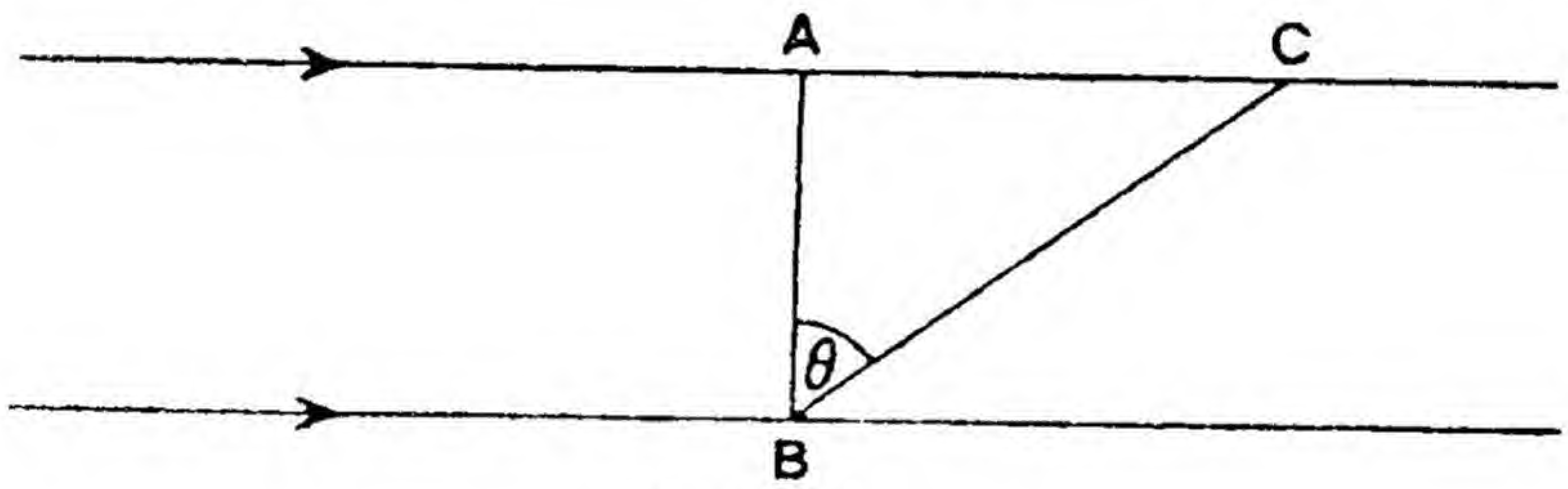


Fig. 144.

all the light which falls on it. Also it is possible for two surfaces to have the same illumination and to have widely differing brightnesses due to a difference in reflecting powers.

We are now in a position to understand the principle underlying the various methods of finding the luminous intensity of a source of light and the illumination of a surface, the two main determinations in the science of photometry. The reader will be familiar with the fact that all scientific measurements are made by means of the senses, but that the advance of science has limited the use of the senses as far as possible, and, in the majority of cases, restricts it to deciding the coincidences between two lines, as, for example, that between the pointer of a galvanometer and the zero mark of a scale in the measurement of a resistance. In the case of photometry it is necessary to make rather greater demands on the eye. It is required to decide when two surfaces are equally bright ; it is found that the eye can do this to an accuracy of 1 per cent. under the most favourable conditions, which are an illumination from 4 to 40 foot-candles. If the two surfaces are made of the same material, when their brightnesses are equal, so are their illuminations, and it is easy to calculate the ratio of the luminous intensities of the two sources illuminating the two surfaces. Suppose the two point sources of candle-power P_1 and P_2 send

light at the same angle i on to the two surfaces of the prism ABC, the face AB being illuminated only by the source P_1 and the face AC only by the source P_2 (Fig. 145). If the distances d_1 and d_2 of the sources from their respective surfaces are adjusted so that the surfaces look equally bright, then their illuminations are equal. Therefore

$$\frac{P_1 \cos i}{d_1^2} = \frac{P_2 \cos i}{d_2^2}$$

$$\therefore \frac{P_1}{P_2} = \frac{d_1^2}{d_2^2} \quad \ldots \ldots \ldots \quad (48)$$

Hence the ratio of the luminous intensities of the two sources can be calculated when the distances d_1 and d_2 have been measured and, if one of the sources is of known candle-power, that of the other can at once

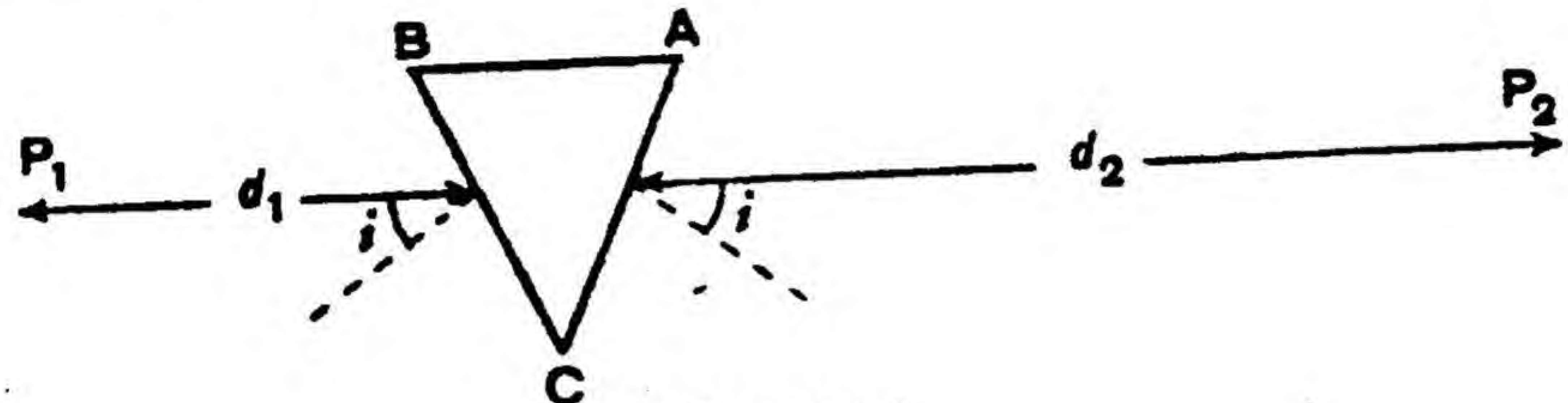


Fig. 145.

be found. The above equation is the fundamental relation from which luminous intensities are calculated with nearly all photometers. Once luminous intensity can be measured, the illumination of any surface can be found by illuminating a similar surface with a lamp of known candle-power and adjusting the distance of the lamp from the surface, so that it appears as bright as the surface under examination. Then the illumination of the artificially lit surface can be calculated from the law of inverse squares and is equal to that of the given surface. The various photometers which are used for measuring luminous intensity differ from one another in the devices adopted for enabling the eye to decide when the two surfaces are equally bright. A number of photometers will now be described.

81. THE GREASE-SPOT PHOTOMETER

This photometer consists of a circular sheet of white paper, which reflects all the light which falls on it and transmits none. It is mounted in a suitable holder and a spot in the centre of it is made translucent to light by dropping a little grease on it. This part transmits a good deal of the light and reflects the rest. The photometer and the two sources are mounted as shown in Fig. 146 and the distances of the two sources from the paper are adjusted, until the spot disappears as viewed from either side of the paper. To avoid having to move the eye about from one side of the paper to the other, two mirrors M_1 and M_2 are placed so that an eye at E can see both sides of the paper at the same time. When the

spot disappears it can be shown that the illumination of each side of the paper is the same, and so the distances d_1 and d_2 of the sources P_1 and P_2 respectively from the screen are measured and the ratio of their luminous intensities P_1 and P_2 is calculated from equation (48).

To see that the illuminations are equal when the spot disappears, let us imagine that 100 lumens fall on unit area of the left-hand side of the paper and that the spot transmits 40 per cent. of the light falling on it and reflects the rest. Each unit area of the paper reflects 100 lumens, but the spot only reflects 60 lumens and, if the illumination on the right-hand side of the paper is zero, this represents the total amount of light sent out by the spot to the left. So, if an observer looks at the paper from the left-hand side, he sees the spot darker than the paper, whereas he will see the opposite from the right-hand side. This is just a simple case of the general rule that the spot always looks darker than the paper when viewed from the side on which the illumination is the greater. Also the spot can only send out as much light as the paper on the left-hand side, if it restores to the reflecting beam in some way the 40 lumens which it has lost by transmission. This can only be done by producing an illumination of 100 lumens on the right-hand side, when the spot will transmit 40 lumens of this and so bring up the amount of light which it emits towards the left to 100 lumens per unit area. This is then equal to the amount of light emitted by unit area of the paper, and so the spot will be indistinguishable from it. The reader can quickly prove for himself that, if the spot disappears when viewed from one side, it will also disappear when viewed from the other.

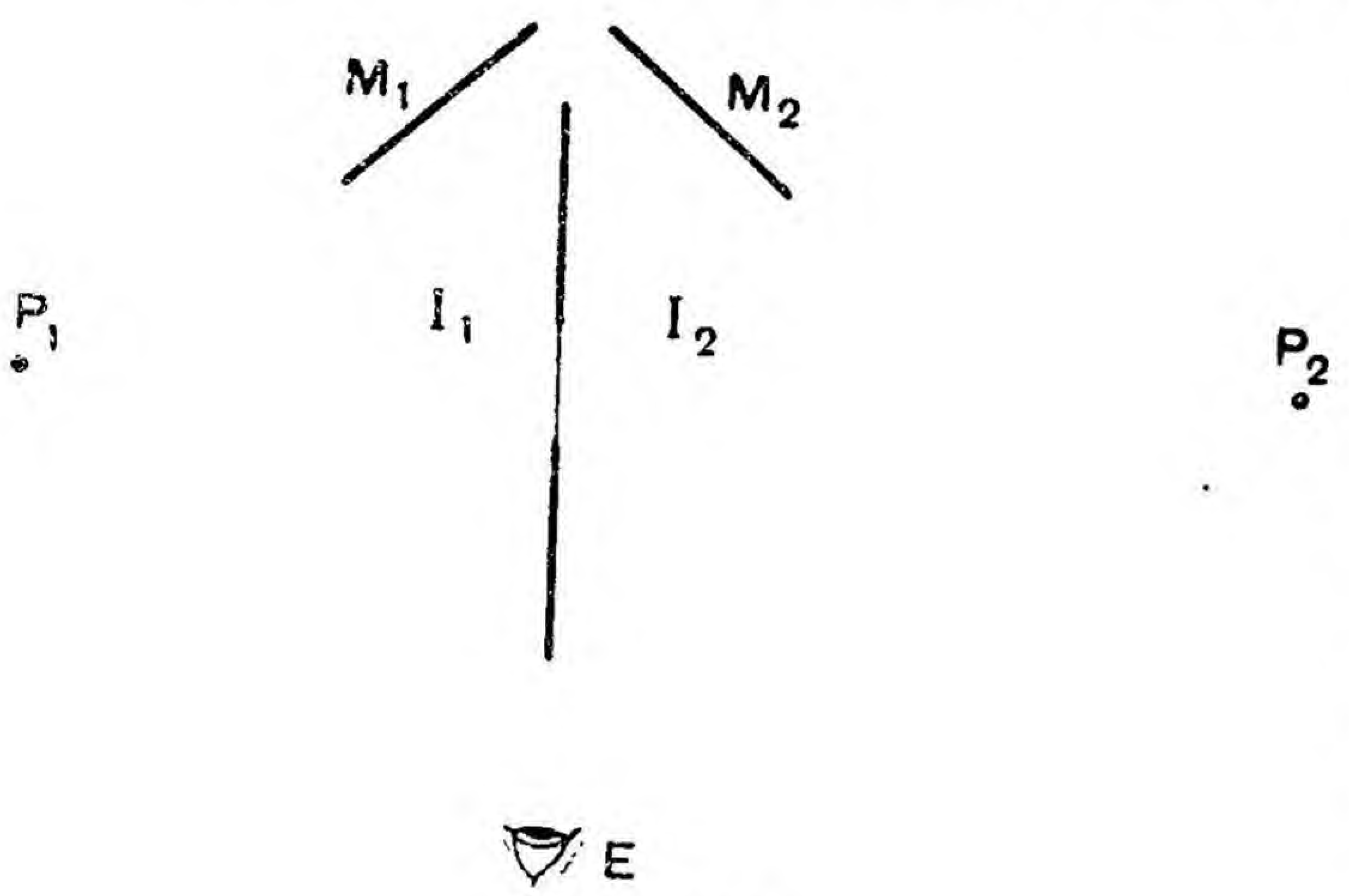


Fig. 146.

As the principle of this grease spot photometer is the same as that of the more accurate Lummer-Brodhun photometer to be described below, we shall discuss the general case. Let the illumination on the two sides be I_1 and I_2 (Fig. 146), then the amount of light reflected diffusely by unit area of the paper is I_1 on the left-hand side, assuming it is a perfect reflector. If the grease spot reflects a fraction r of the light falling on it and transmits the rest, both the reflection and transmission being diffuse, then the amount of light sent towards the left by the spot is $rI_1+(1-r)I_2$. Hence the spot disappears if

$$I_1=rI_1+(1-r)I_2$$

that is, if

$$I_1=I_2$$

or if the illuminations on the two sides are equal. If this is so, then it is clear that the amount of light sent towards the right by the paper and the spot is also the same, and so the spot will disappear on the other side also.

Now it happens in practice that the spot does not disappear simultaneously on both sides; if it disappears on the left-hand side, say, it looks darker than the paper on the right-hand side. This would seem to indicate that the illumination is the greater on the right-hand side. Its disappearance on the other side is accounted for by the fact that the spot reflects some of the light, *absorbs some*, and transmits the rest. On account of this absorption, what the reflected beam lacks on the left-hand side cannot be made up by that transmitted from an equal illumination on the other side, but must be transmitted from a slightly greater illumination. Consequently, in practice, the two sources of light are adjusted so that there is *equal contrast* between the spot and the paper on the two sides. For this reason it is especially desirable to have the two mirrors M_1 and M_2, so that the two sides of the paper can be observed at the same time.

82. THE JOLY PHOTOMETER

It is not easy to decide when the contrast between the spot and paper is the same in the grease-spot photometer; this means that one of the sources may be moved by some 5 per cent. of its distance from the paper before any noticeable change in the contrast is produced as between the two sides. So Professor Joly, of Trinity College, Dublin, designed a new form of photometer in which he hoped that equality of illumination could be more accurately judged.

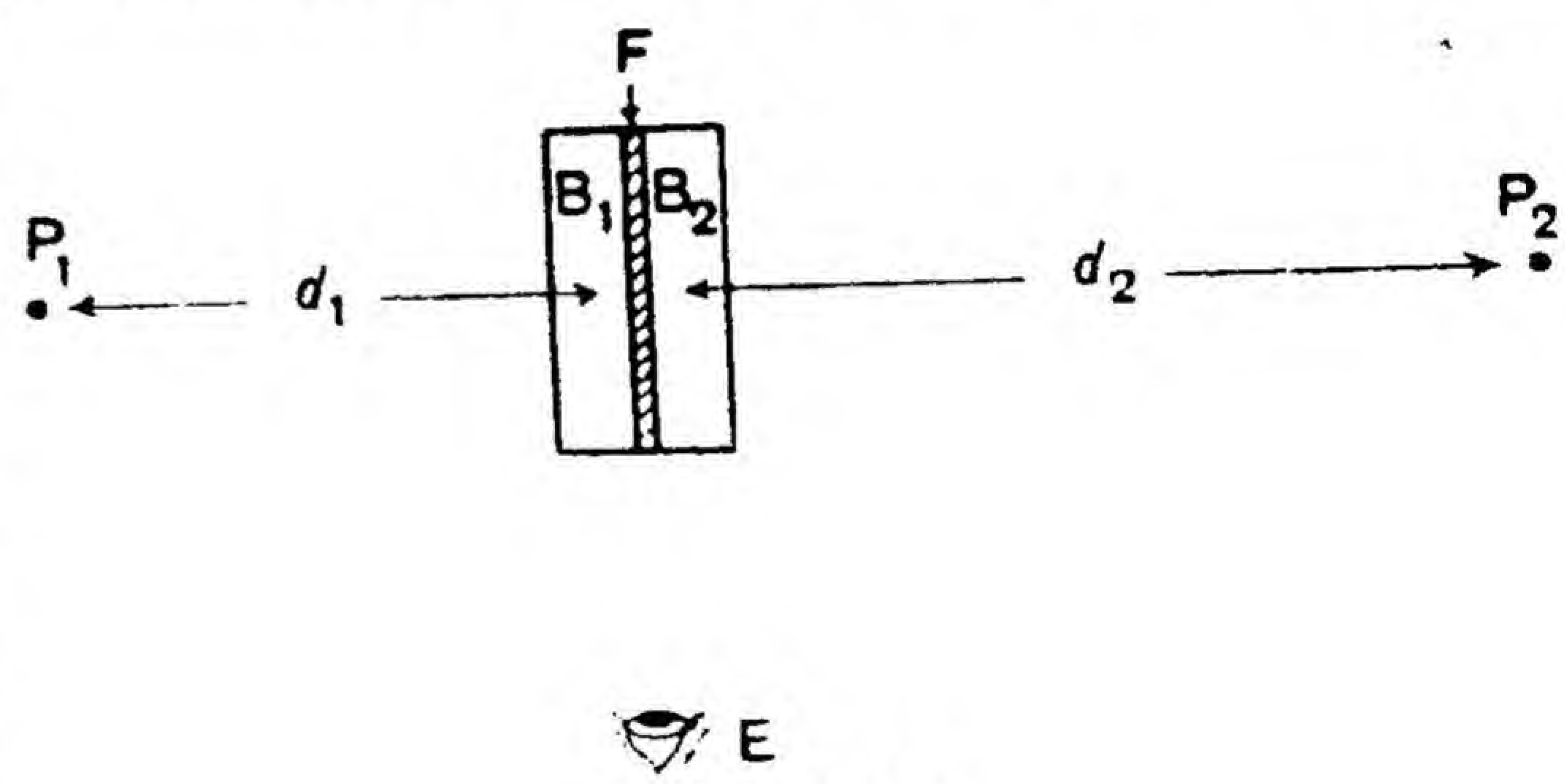


Fig. 147.

It consists of two paraffin wax blocks B_1 and B_2, as nearly as possible identical, separated by a piece of tinfoil F (Fig. 147). The blocks are mounted with the line joining the sources P_1 and P_2 normal to the surfaces of the blocks and so that the light from the source P_1 illuminates only B_1 and that from P_2 only B_2. The distances d_1 and d_2 are adjusted so that the two blocks look equally bright when viewed by an eye at E; d_1 and d_2 are then measured and the ratio of the luminous intensities is calculated from equation (48). It is advisable to repeat the experiment with B_1 and B_2 reversed in case they are not identical in thickness or reflecting power.

83. THE LUMMER-BRODHUN PHOTOMETER

One of the difficulties in trying to decide if two surfaces are equally bright is that the surfaces are usually separated by a line of small but

finite width; the observer always has the feeling that he could make a more accurate comparison if the two surfaces actually touched one another. Lummer and Brodhun succeeded in designing a photometer incorporating this feature. The two sources P_1 and P_2 (Fig. 148), whose luminous intensities are to be compared, illuminate the two sides

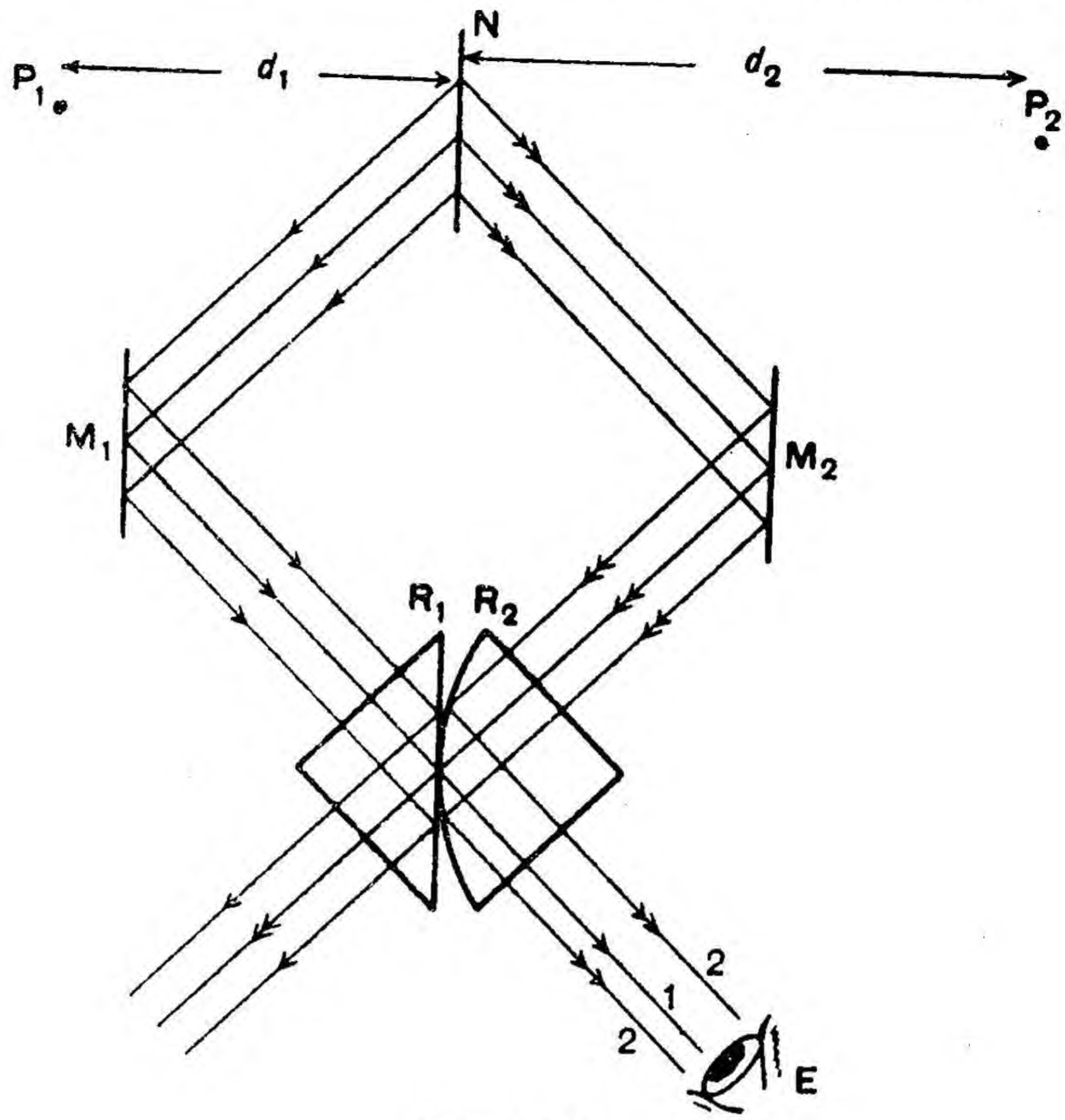


Fig. 148.

respectively of the screen N, which is painted with a very good reflector such as magnesium carbonate. Some of the light reflected from the left-hand side of this screen is reflected by the plane mirror M_1 towards the two prisms R_1 and R_2. R_1 is an isosceles right-angled prism with a flat base, while R_2 is of similar design only with a slightly convex base, and so the two prisms are only in contact for a small circular patch in the middle of their bases. Therefore the inner rays from the mirror M_1 pass straight through the two prisms and come out and fall into the eye at E, while the rays on the margin of the beam are totally reflected by R_1, since they strike its base at a point where it is separated from the base of R_2 by an air film. In the same way the inner rays of the beam reflected from M_2, which originated in the right-hand side of N, pass straight through the prism combination, while the outer rays are totally reflected and enter the eye. The field of view presented to the observer is shown in Fig. 149 and consists of an inner circle 1, which is really an image of a small portion of the left-hand side of N formed by reflection in M_1, and an

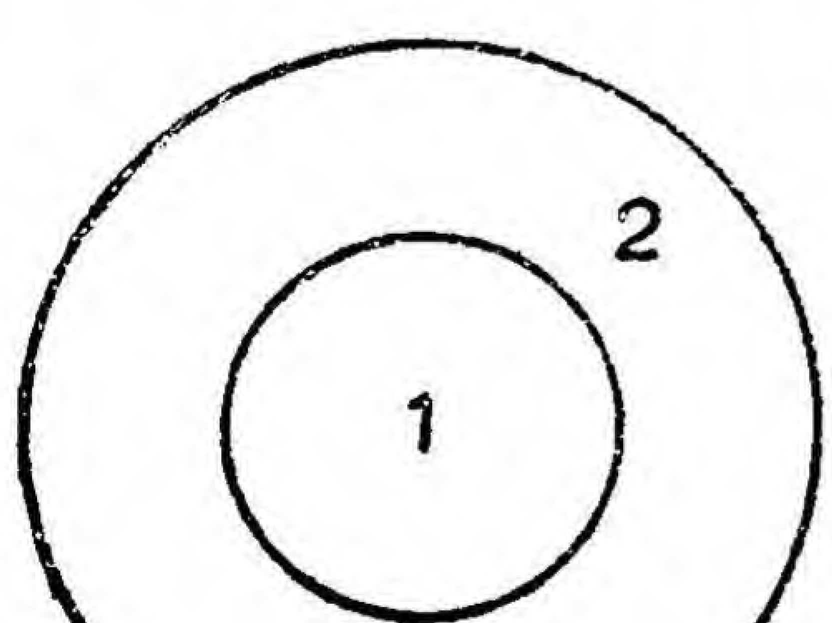


Fig. 149.

outer zone 2, which is an image of a portion of the right-hand side of N formed by reflection in M_2 and total reflection in R_2. There is practically no line of demarcation between these two zones, since the transition from total reflection to transmission at the bases of R_1 and R_2 takes place in a very short distance. The observer views the field and adjusts the distances d_1 and d_2 of P_1 and P_2 respectively from the screen N, until he cannot distinguish the zones 1 and 2 from one another. The brightnesses and therefore the illuminations of the two sides of N are then equal and the ratio of the luminous intensities P_1 and P_2 of the two sources is then calculated from equation (48) in the usual way.

The Lummer-Brodhun photometer is the most accurate that has so far been made, which means that a smaller percentage change in one of the distances is needed to produce a noticeable difference in the brightnesses of the two surfaces in this photometer than in any other. But it suffers from the disadvantage that it cannot give accurate results in the case of sources of different colour; this means that a given observer finds that he can move one source a comparatively large distance without revealing any appreciable difference in the brightnesses of the two illuminated surfaces. Also, if one observer sets the photometer so that the two differently coloured parts of the field appear to be of equal brightness to him, they will appear to be of quite different brightness to another observer. This inaccuracy is not a defect of the photometer itself so much as one of the human eye, which is unable to make an accurate judgment of the equality of brightness of two different colours. It can be overcome by spreading out each source into its spectrum and comparing the luminous intensities of the two sources wave-length by wave-length and then integrating up the whole effect. Details of this method are beyond the scope of this book and the reader should consult Hardy and Perrin's "Principles of Optics" for an account of this technique, known as spectro-photometry. But a simple method of comparing the brightness of two different colours is by the Flicker photometer, which will now be described.

84. THE FLICKER PHOTOMETER

This photometer can be used both to compare the luminous intensities of sources of the same colour as well as those of different colour. It consists of a white screen N (Fig. 150) and a Maltese cross C, which can be rotated at a suitable speed about an axis normal to its plane by an electric motor. The screen and cross are each painted white and are arranged to be at the same angle to the light incident on them and to the tube T down which they are observed by an eye at E. The screen N is illuminated by one of the sources P_1, while the cross is illuminated by the other source P_2, and if the cross is rotated the eye at E has presented to it at regular intervals first the screen N and then the cross C. If the brightnesses, and therefore the illuminations, of the two surfaces are different, then a

sensation of flickering is experienced which can be made to disappear by adjusting the distances of the two sources from the surfaces which they illuminate. When the flickering vanishes, the illuminations of the two surfaces are the same and the distances d_1 and d_2 of P_1 and P_2 from the screen and the cross respectively are measured and the ratio of the luminous intensities of the two sources is calculated from equation (48). It is important to adjust the speed of rotation of the cross so that the eye is most sensitive to flicker, the speed being different for different illuminations of the screen.

If the luminous intensities of two different colours are being compared, the cross is set rotating and a uniform colour is produced ; superimposed on this is a sensation of flicker, which can be made to disappear by a suitable adjustment of the distances of the sources from their respective screens. When this is the case, the brightnesses of the two different coloured surfaces are said to be the same. Hence their illuminations are equal and so the luminous intensities of the two sources can be calculated

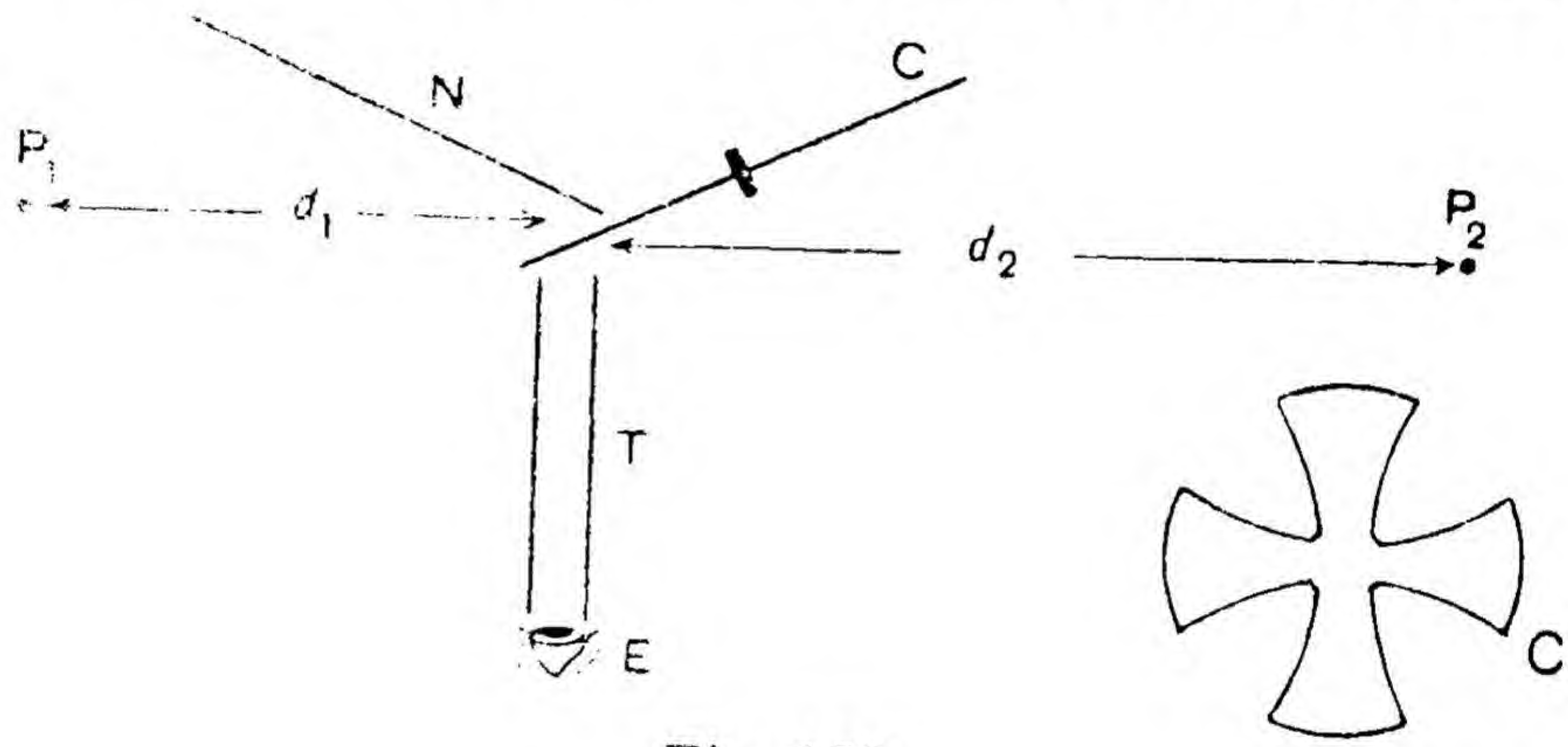


Fig. 150.

in the usual way. It must be emphasised that this statement that the brightnesses of the differently coloured surfaces are the same when the flicker disappears is an arbitrary one, but it is found that the results obtained on this assumption agree within the limits of experimental error with those obtained by the more fundamental method referred to in the previous article.

Certain precautions must be observed if accurate results are to be obtained with any photometer. It is necessary to use sources of accurately the same colour except with the Flicker photometer, otherwise it is not possible to make an accurate setting for equal brightnesses of the two illuminated surfaces. When the brightnesses of the two surfaces become nearly equal, the eye seems to be very sensitive to differences in colour, and, if the two surfaces are not identical in colour, it is quite impossible to fix the position of the two sources for which the brightnesses of the two surfaces are the same. It is possible to move one of the sources through quite a considerable distance while the other remains fixed, before the eye can be certain that one surface is fainter than the other. Secondly, the distances of the sources from the surfaces should be so adjusted that

the illumination lies between 4 and 40 foot-candles, since the eye is most sensitive to changes of brightness over this range of illumination. It has been shown experimentally that it can detect a difference of illumination of 1 per cent. in two surfaces within this range of illuminations, and we can see the bearing of this on the accuracy of the final results. Let us consider the grease-spot photometer, in which the sources are adjusted so that the grease spot disappears. The reader will recall that this neglects absorption, but this is not pertinent to our present discussion. Using the same notation as when discussing the grease-spot photometer, the spot disappears when

$$I_1 = rI_1 + (1-r)I_2$$

Really it disappears when

$$rI_1 + (1-r)I_2 = I_1 \pm \tfrac{1}{100}I_1$$

or

$$(1-r)I_2 = (1-r)I_1 \pm \tfrac{1}{100}I_1$$

that is, when

$$I_2 = I_1 \pm \frac{I_1}{100(1-r)}$$

Therefore if $r \to 1$, even when the spot does disappear, there is a considerable uncertainty in the equality of I_1 and I_2. Now the Lummer-Brodhun photometer is really a grease-spot photometer, in which the circular patch of the bases of the two prisms which are in optical contact replaces the grease spot. For this spot $r=0$, and so the uncertainty in the equality of I_1 and I_2 is a minimum. This is another reason why the Lummer-Brodhun photometer is one of the most accurate of all photometers. Finally, it is necessary to eliminate all stray light; each surface must be illuminated by the appropriate source and *must receive no light from any other place.* If stray light is not excluded, it makes it more difficult to decide when the two surfaces are equally bright. For example, let the two surfaces be illuminated with an intensity of 100 foot-candles from their respective sources. Then, if one source is moved so as to alter the illumination of its surface to 99 foot-candles, the observer will just detect a difference in brightness of the two surfaces. But, if the two surfaces have stray light of intensity 50 foot-candles in addition to the 100 foot-candles from their respective sources, it is necessary to move the source the further distance needed to produce an illumination on its surface of 98·5 foot-candles, making 148·5 foot-candles in all, before the eye can detect any difference in brightness between the two surfaces. So the room in which photometry is carried out should have the walls painted black; screens must be placed so that any of the light striking the stands on which the photometer is mounted cannot get to either surface; and the apparatus must be so disposed that the light from one source illuminates only one surface and sends no light to the other surface either directly, or indirectly via the walls of the room, or stands, or other apparatus. If these precautions are observed it is possible with practice

to obtain results consistent to 1 per cent. or perhaps $\frac{1}{2}$ per cent., but beginners will be considerably less accurate than this, as their eyes are untrained in detecting small differences of brightness.

85. MEAN SPHERICAL CANDLE POWER

Any ordinary photometer only operates by using the light which actually falls on the screen or object to be illuminated; that is, it only uses a small portion of the total light emitted by the source and so it really measures the luminous intensity in a certain direction. It remains to be seen if the luminous intensity is the same in all directions, and this can only be settled by an experiment of the following kind. A lamp is placed in the position shown in the upper diagram in Fig. 151 with respect to the screen N of a Lummer-Brodhun photometer and the luminous intensity is measured in the usual way. It is then turned through 90° so as to be in the position shown in the lower diagram of Fig. 151 and the luminous intensity is again measured. It is found with most of the common sources of light, such as gas mantles and electric light bulbs, that the luminous intensity in these two positions is quite different, showing that the candle power of a source of light does vary with direction. This variation has to be investigated experimentally by the method outlined above, the experiment being confined to one plane through the axis of symmetry which most sources of light possess. When the candle-power in a suitable range of directions in this plane has been measured, the results are plotted on a diagram illustrated in Fig. 152, in which the candle-power in a direction making an angle of 30° with the horizontal in a downwards direction is given by the length of the line OA. Such a curve is called the **polar curve** of the source of light and the diagram shows a typical curve for an old-fashioned filament lamp. It will be seen that the candle power in a direction perpendicular to the length of the filament is considerably greater than that in the same direction as the length of the filament and is greater in directions below the bulb than above it.

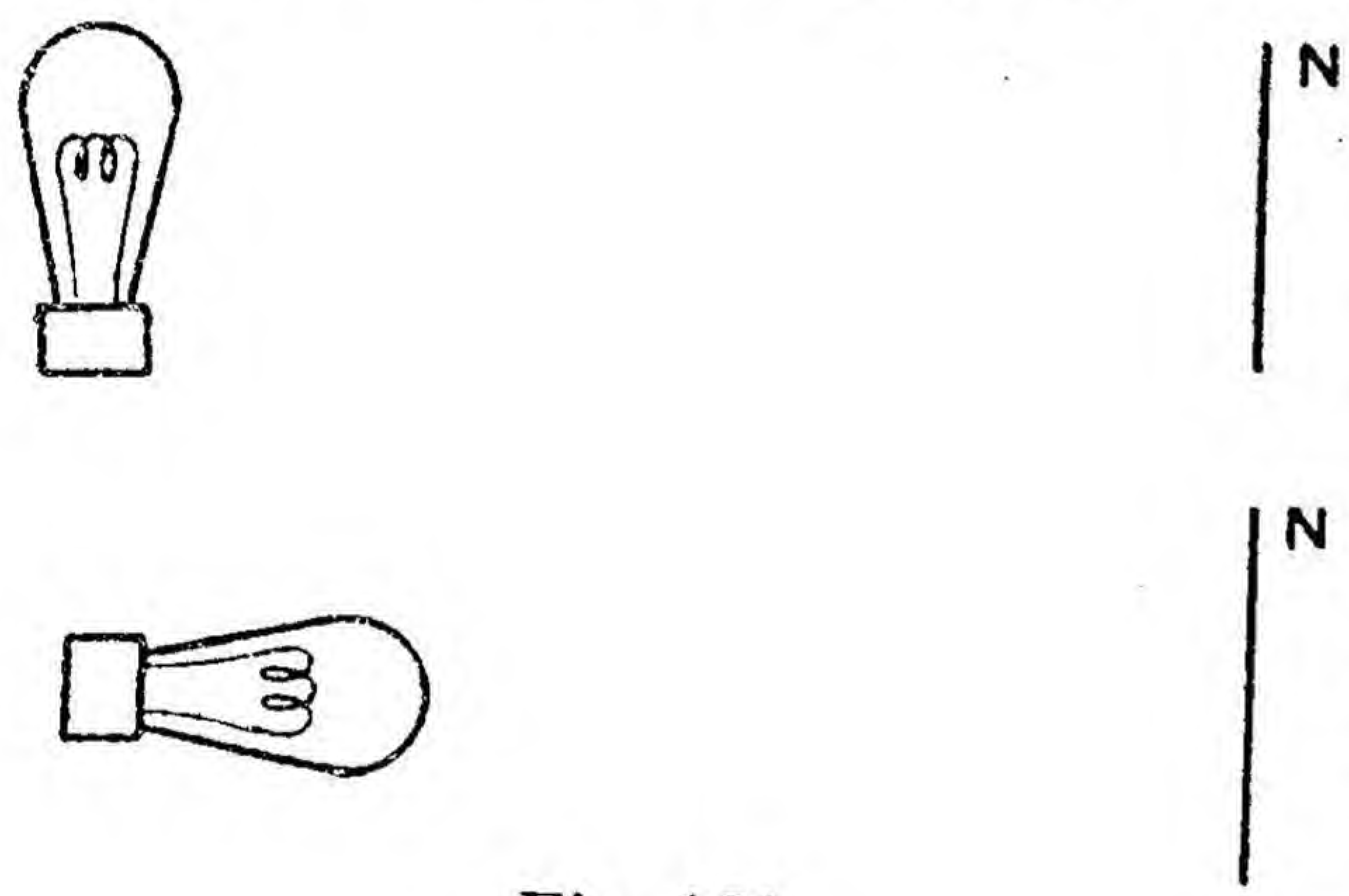


Fig. 151.

This result makes it necessary to revise our definition of the luminous intensity of a source. If dL is the amount of light falling in 1 second in lumens on an area $d\alpha$ placed 1 cm. from the source so that the light falls normally on it, the luminous intensity P in the given direction is given by

$$P = \frac{dL}{d\alpha}$$

Once the polar curve has been found it is possible to calculate the **mean**

spherical candle-power of the source, which is equal to the total amount of light emitted by the source in 1 second in lumens divided by 4π. The polar curve is usually found to be the same for any plane passing through the axis of symmetry of the source and, assuming this to be the case, the procedure is as follows. The polar curve is divided into n sectors each covering an angle of $\frac{2\pi}{n}$ radians, and a sphere of unit radius is described about the source as centre. If the candle-power in the sector making an angle θ with the horizontal is P, then the light emitted in this direction

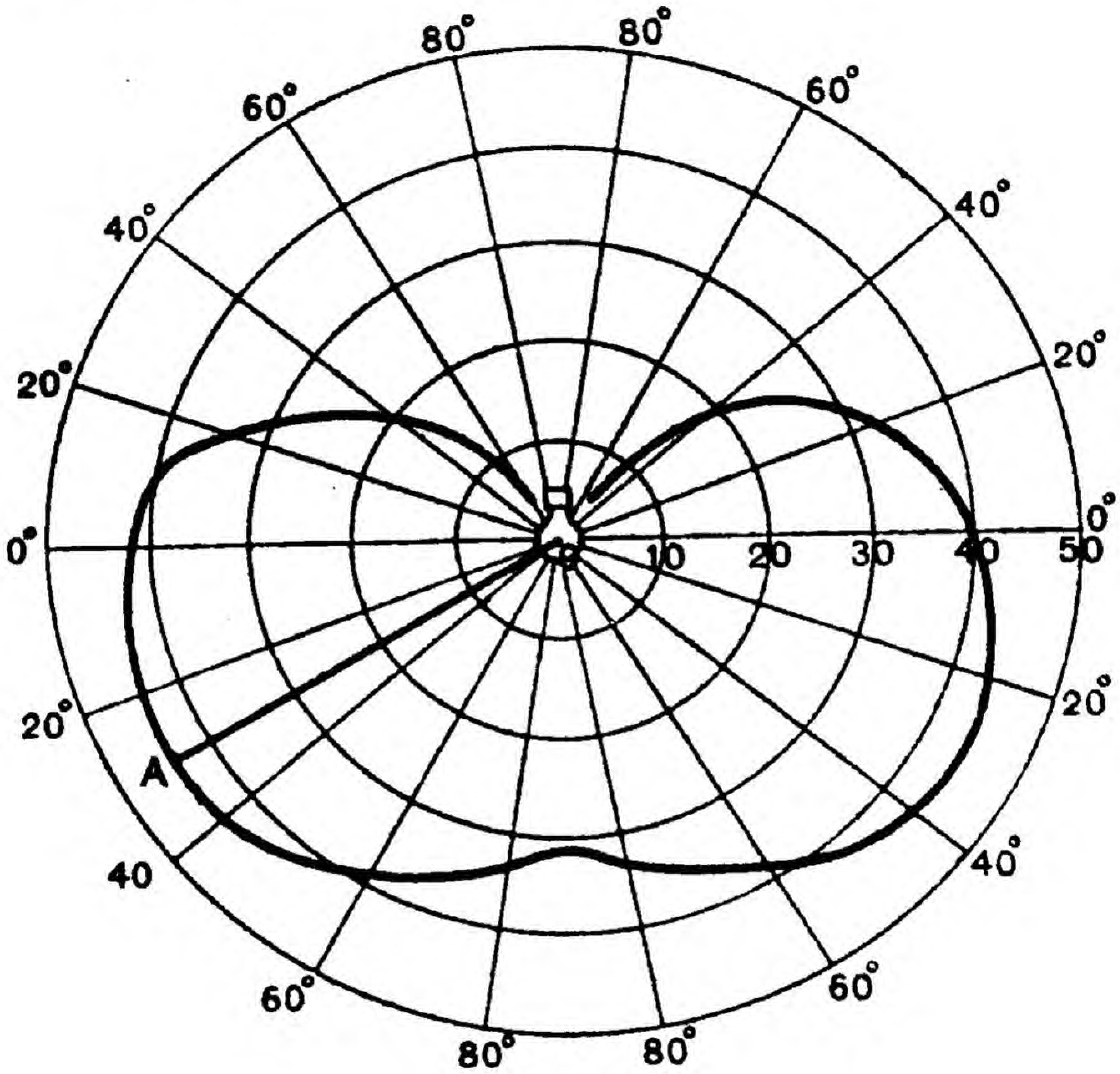


Fig. 152.

falls on a zone of the sphere formed by rotating the lines bounding the sector about the axis of symmetry of the source, which is assumed vertical. Therefore the length of the zone is $2\pi \cos\theta$ and its breadth is $\frac{2\pi}{n}$. The amount of light falling in 1 second on this zone is

$$P \times 2\pi \cos\theta \times \frac{2\pi}{n} \quad \text{or} \quad \frac{4\pi^2}{n} . P \cos\theta$$

Thus the total amount of light emitted by the source in 1 second is

$$\frac{4\pi^2}{n} \sum P \cos\theta \text{ lumens}$$

the summation being taken over the n zones into which the sphere has been divided. This is the quantity which is really of interest in con-

nection with a source of light, the mean spherical candle-power being found by dividing it by 4π.

This operation is somewhat tedious and it can be replaced by a direct experimental method, once the mean spherical candle-power of a source has been found by the above method. The source of light P_1 whose mean spherical candle-power is required is placed inside an **integrating sphere** (Fig. 153), which is a large sphere some 3 to 9 feet in diameter, the inside of which is painted white and which is provided with a small hole A out of which light can escape. The surface of the sphere is a perfect diffuse reflector and as a screen N is placed so as to prevent any light from escaping directly from the source P_1 through the hole A, the light which issues from it is proportional to the total quantity of light emitted in unit time by the source. The reason for this is that light from almost every part of the inside of the sphere contributes to the light escaping from the hole and so light emitted in every direction from the source is included. Thus the sphere is called an integrating sphere because it performs experimentally the integration described in the previous paragraph. The luminous intensity of the hole A regarded as a source of light is compared with that of another source S using a Lummer-Brodhun photometer, or some other suitable type, and then the source of required candle-power is replaced by another P_2, whose mean spherical candle-power has been found by the method described above. The luminous intensity of the hole A is again compared with that of the same source S and the ratio of the luminous intensities of P_1 and P_2 can now be calculated. It is equal to the ratio of the corresponding mean spherical candle-powers. As that of P_2 is known, the mean spherical candle-power of P_1 can be calculated.

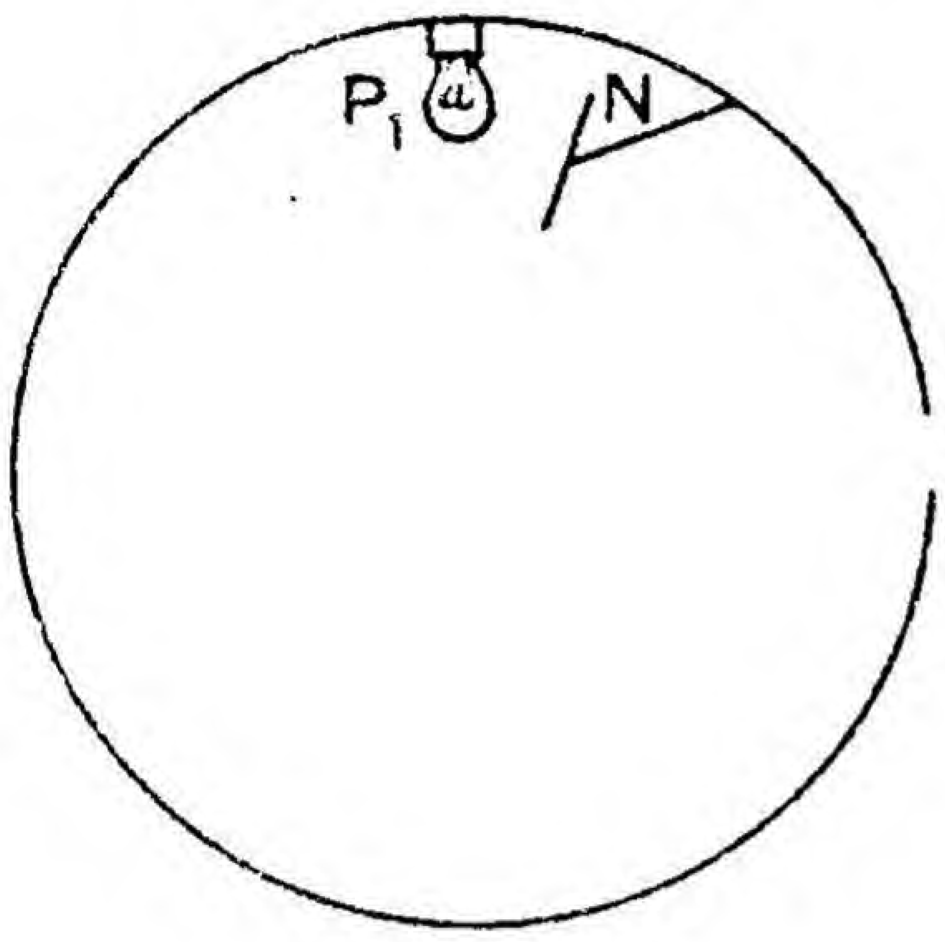


Fig. 153.

86. THE MEASUREMENT OF ILLUMINATION

The reader will recall that the illumination of a surface is the amount of light energy falling on unit area of it in unit time and is measured either in lumens per square foot or in foot-candles. We have already emphasised that the measurement of illumination is essential if the lighting of factories and streets is to be improved, for both the illumination regarded by normal persons as adequate and the present illumination must be known in order to decide whether improvement is needed and, if so, to what extent. It is hardly necessary to add that it is not much use to the illumination engineer to know that the illumination at a particular point in a factory is bad and that it must be improved. He must be able to obtain numerical measures of these quantities if any real progress is to be made. There are two classes of illumination meters; the first

produces a surface of known illumination, which is judged by the eye to be of the same brightness as the surface under test ; the illumination of that surface is then equal to the known illumination, if the two surfaces have the same reflecting power, as can easily be arranged. The second class consists of those which convert the light falling on the surface into an electric current and determine the illumination from the strength of the current so produced.

The Weber Illumination Meter (Fig. 154) is an example of the first class and consists of a Lummer-Brodhun photometer head L, which

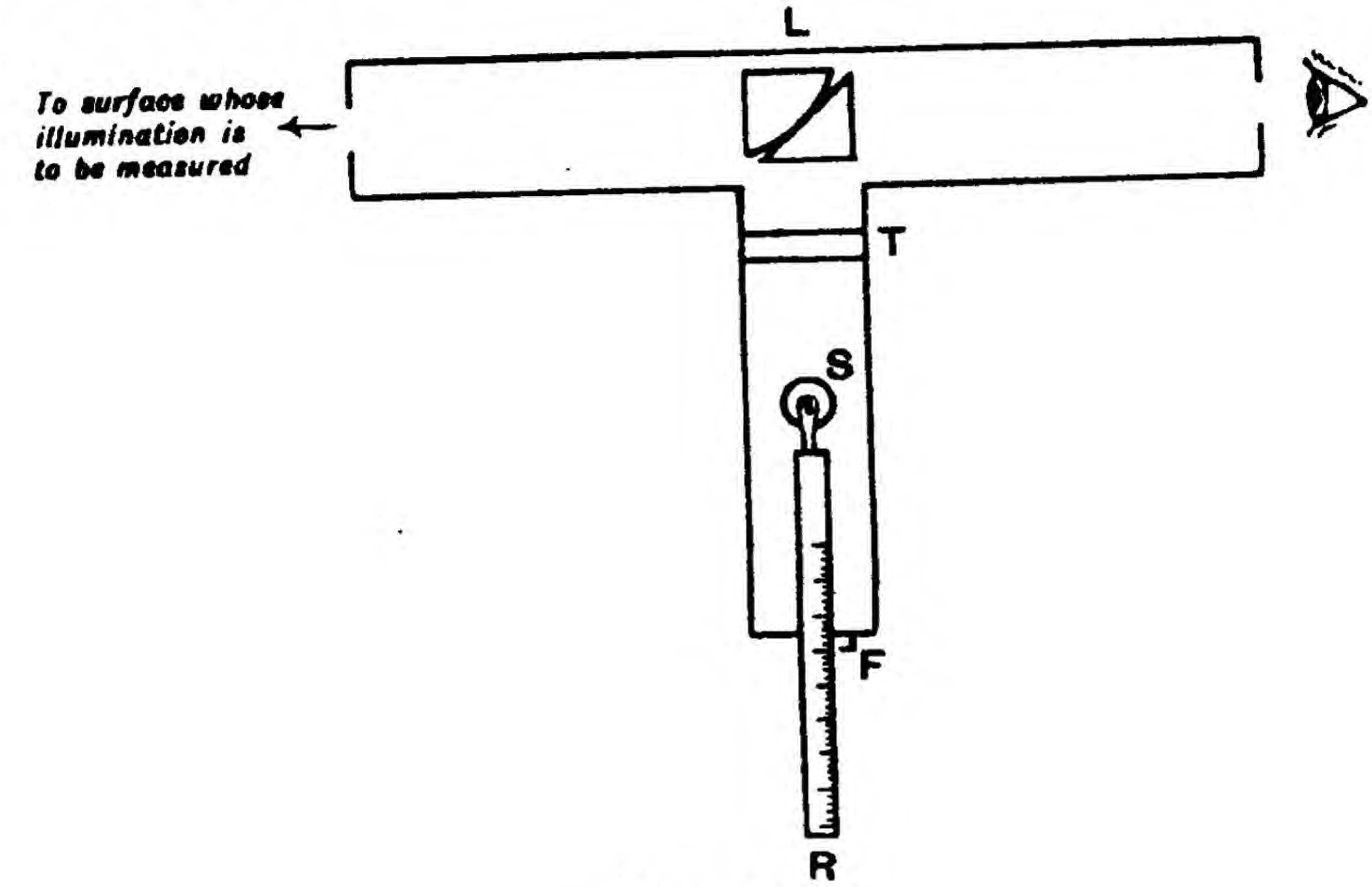


Fig. 154.

receives light from the surface to be tested and also from the translucent plate T, which is illuminated by the lamp S. The distance of this lamp from the plate can be altered so as to produce a variety of illuminations on it, and the rod R on which it is mounted has a scale which is calibrated in foot-candles in the following way. The unit shown in Fig. 155 is placed on the test plate P and the current through the lamp M is adjusted to a definite value, from which its candle-power is known, and hence the illumination of the plate I can be calculated from this candle-power and the distance from the lamp to the plate. Let us suppose the illumination is adjusted to be 1 foot-candle. The illumination meter is then sighted at the plate P through the hole H, and the distance of S from the plate T is adjusted until the two halves of the photometer field appear equally bright, and the point on the scale R opposite the fiducial mark F is marked 1 foot-candle. This process is then repeated at suitable illuminations over the range for which the meter is

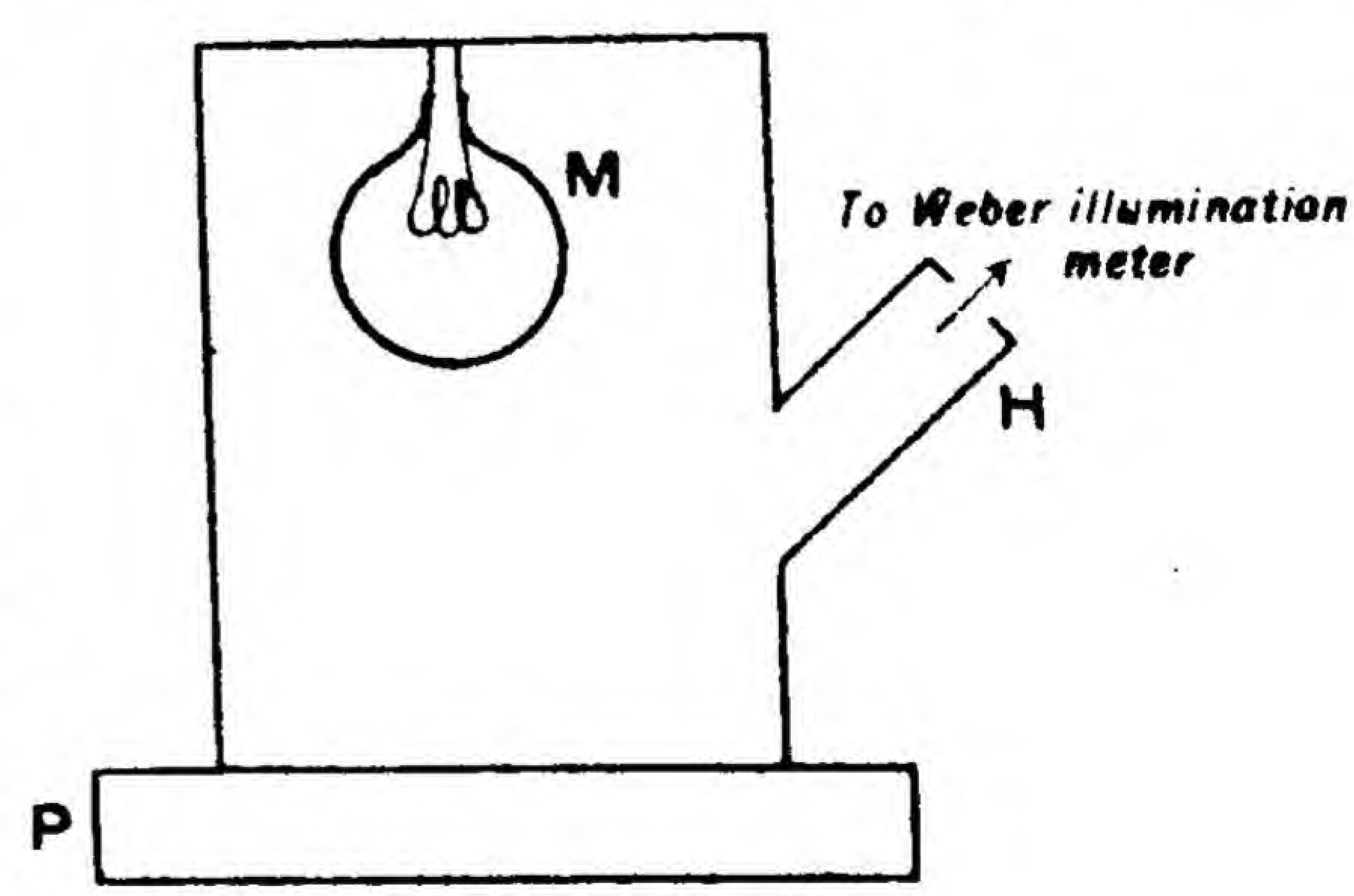


Fig. 155.

to be used, and the illumination of any surface is then measured by placing the test plate P on the surface and sighting on it with the illumination meter. The distance of the lamp S from the plate T is adjusted for equal brightnesses of the two halves of the photometer field, when the reading of the scale R gives the illumination of the surface in foot-candles. It must be emphasised that this instrument measures the illumination of the surface, which refers to the amount of light falling *on* it and not to the amount given *off* by the surface. Hence there is no objection to covering the surface with a test plate, which may be of a different material, since this does not affect the amount falling on the surface. The reader will appreciate that the test plate must always be used, since the calibration was carried out with that plate and it is essential that the same fraction of the light falling on the surface should be sent to the meter in actual use as in calibration.

The other type of illumination meter is based on the photo-electric effect, which will be discussed more fully in Chapter XVII. It is well known that, when light of a suitable wave-length falls on a metal plate, electrons are emitted and that the number of electrons emitted per unit area per unit time is proportional to the light energy falling on unit area of the plate in unit time. The photo-electric illumination meter does not make use of a metal, but of the copper-cuprous oxide surface which is used in metal rectifiers. This is prepared by heating pure copper in air, when it becomes coated with a layer of cuprous oxide covered by a thicker layer of cupric oxide. This outer black layer is either removed mechanically or dissolved off, and it is found that the resistance of the surface to electrons flowing towards the copper is several thousand times greater than that to electrons flowing away from the copper. If light is sent on to the cuprous oxide surface, it behaves like a metal and emits electrons, and much larger currents are obtained in this way than with a metal surface alone. It is usual to coat the cuprous oxide surface with a very thin layer of gold or aluminium (Fig. 156), but the light is able to penetrate this to reach the cuprous oxide layer. The current produced is proportional to the illumination on the surface and is of the order of 10^{-5} amps per lumen. The illumination meter simply consists of a copper-cuprous oxide surface joined to a sensitive moving coil galvanometer, which is calibrated directly in foot-candles using a lamp of variable known candle-power as described in the case of the Weber Illumination Meter. The photo-electrical lumination meter is rapidly replacing the other type, since its response to different colours is very similar to that of the eye and it is a pointer-reading instrument demanding no visual skill. It may be mentioned incidentally that it is admirably adapted for performing experiments in the laboratory, for it is easy to use and accurate results

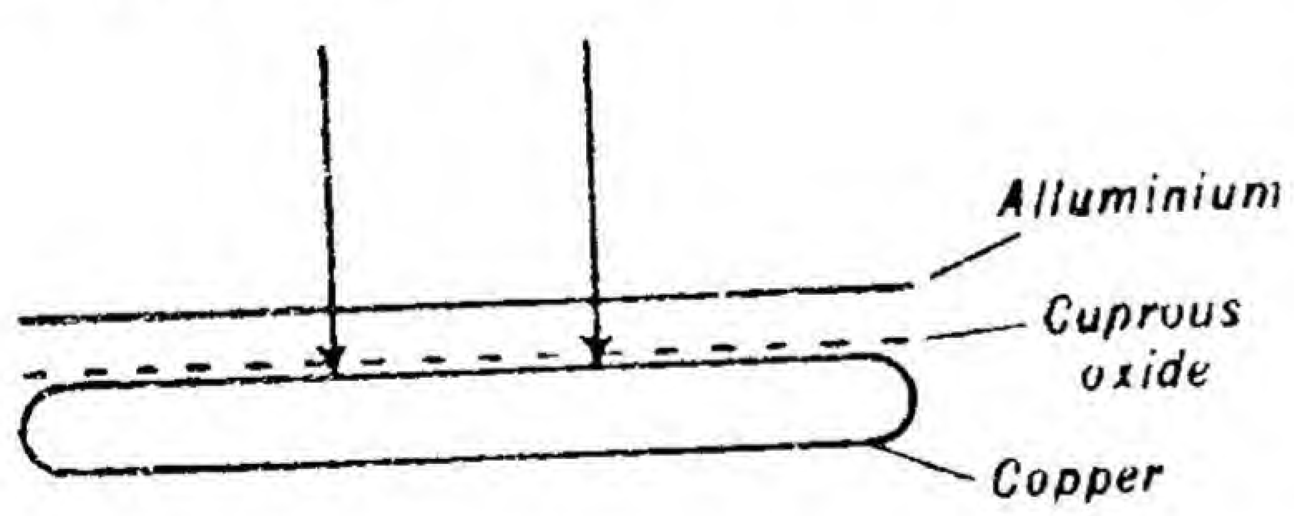


Fig. 156.

can be quickly obtained. For example, it can be used to see to what accuracy a given source obeys the law of inverse squares by measuring the illumination produced by the source with the meter at various distances away from the source. It also gives a simple method of measuring the luminous intensity of a source; if the source produces an illumination of 10 foot-candles at a distance of 2·5 feet, then the candle-power is $10 \times 2{\cdot}5^2$ or 62·5 candle-power. It can also be used for measuring the way in which the candle-power of a source varies with direction and so for obtaining the polar curve of the source. But its chief practical use is as an illumination meter, and it is being used both in illumination engineering and is also being sold to photographers as an exposure meter.

87. MODERN ILLUMINATION ENGINEERING

We shall now consider briefly how the ability to measure the luminous intensity of a source and the illumination of a surface has been used to improve the lighting of buildings, factories, and streets. This is now done scientifically and the name given to this branch of knowledge is Illumination Engineering. The first problem to be settled is the suitable illumination for various types of work, and this has been done by measuring the illumination at which the work is carried out and varying it until the average worker expressed himself satisfied that he could work comfortably. Tests were carried out with a large number of workers in each occupation and an average value of the comfortable illumination for each occupation has been obtained. Both the Electric Lamp Manufacturers' Association and the Illuminating Engineering Society have made tests and the type of results which they obtained is shown in Table 11.

When the comfortable illumination for various occupations has been settled, it is possible to design scientifically the lighting system in a room so as to provide the recommended illumination. Let us imagine that we

TABLE 11

Occupation or Type of Room.	Recommended Illumination in Foot-Candles.
General Stores.	2–4
Classrooms in Schools.	5–10
Ordinary Shops.	10–15
Tailor's shop; operating theatre; billiard table.	20
Type-setting and engraving.	35–50
Operating-table.	100

wish to instal electric lights in the lecture theatre of a University so as to produce an illumination of 10 foot-candles all over the benches. If the room is 80 ft. long and 40 ft. wide, an illumination of 10 foot-candles or 10 lumens per sq. ft. must be produced all over an area of 3200 sq. ft.,

so the total amount of light falling on the benches and floor per second is 32,000 lumens. But the lights installed must emit light at a greater rate than this, since some of the light emitted is lost by absorption in passing through the fittings and some by reflection off the walls and ceiling. So we must find the **Utilisation Factor** of the room, which is the ratio of the amount of light utilized to the amount emitted. It depends on the dimensions of the room, the height of the benches above the floor, the reflecting power of the walls and ceiling, and on whether the light comes directly from the lamps to the benches or is thrown up to the ceiling and then reflected down. The value can be obtained from tables; let us suppose it to be $\frac{1}{2}$ in this case. Then light must be emitted at the rate of 64,000 lumens. This is quite satisfactory at the beginning when the lamps and fittings are new, but the efficiency of the lamps decreases with age, the reflectors, walls, and ceiling get dirty, and so some allowance must be made for depreciation. It is usual to multiply the rate of light emission by the **Depreciation Factor**, which is usually taken as 1·4 for interior lighting. So light must be emitted at the rate of $64{,}000 \times 1{\cdot}4 = 90{,}000$ lumens to maintain an illumination of 10 foot-candles under average working conditions. If 100-watt coiled lamps of efficiency 15 lumens per watt are used, we shall need $\dfrac{90{,}000}{1500}$ or 60 lamps in all to light the room. They will naturally be arranged regularly about the room, so that the illumination may be uniform over all the lecture benches and not bright in some places and dark in between. Many readers must have noticed the great improvement which has taken place in recent years in the lighting of hotel lounges, lecture theatres in Universities, and public halls, and this has been largely due to the application of these principles.

Precisely the same ideas have been followed both in improving the lighting of the roads, a most important problem in view of the 6000 people who are killed each year in road accidents, and in designing systems for floodlighting buildings, a feature which can be used both for advertising and for revealing in a new light the aesthetic features of cathedrals and monuments. In road lighting, for example, a minimum illumination of 2 foot-candles at the point midway between successive lights is recommended on roads carrying important traffic. In floodlighting 1·25 to 3·5 lumens per sq. ft. is recommended in order to make the building stand out above the surroundings.

There is still another factor which has contributed to the improvement in domestic and industrial lighting in recent years, the increase in efficiency of sources of light. It is one thing to say to a manufacturer that 10 foot-candles is the appropriate illumination for his factory, or to a public authority that 2 foot-candles is the minimum illumination for a certain main road passing through their area; but it is quite another thing to persuade them to adopt the lighting system needed to produce the given

illumination. The manufacturer will reply at once that he quite agrees that it is desirable to increase the illumination in his factory, but, in these days of keen competition, how can he afford the extra cost ? Sometimes the reply is that the extra cost is negligible compared to the saving due to increased illumination ! For example, type setting can be done as efficiently in an illumination of 20 foot-candles as in daylight, but at 2 foot-candles the speed of the work drops to three-quarters and the mistakes made are doubled. But the cost of increasing the illumination from 2 to 20 foot-candles is only one-twentieth of the money saved by the resulting increased speed and decrease in errors. What reply can be made to the public authority which objects to the increased cost of road lighting to give illuminations of 2 foot-candles ? Apart from the contention that the safety of human life is not to be measured against increased cost, it can be urged that the improvements which have been made in the luminous efficiency of electric lamps do make it possible to provide road lighting up to the above standard of illumination without an undue increase in cost.

88. THE DEVELOPMENT OF ELECTRIC LAMPS

The efficiency of an electric lamp is usually measured in lumens per watt, since this gives the intensity of the sensation received by the eye for a given power input. It can also be given as the ratio of the rate of emission of visible radiation in watts to the rate of supply of electrical energy in watts expressed as a percentage. The efficiencies of a number of electric lamps are given in Table 12, and it is interesting to sketch the history of the development of the electric lamp to see the progress that has been made in increasing its efficiency. The carbon arc lamp was first invented by Davy in 1810 and has been used in street lighting, but it needs a good deal of attention and is not reliable enough. The carbon filament lamp was invented by Edison in 1879, but its low efficiency of 3·5 lumens per watt is due to the fact that it cannot be run at a high temperature, otherwise its life would be so short. Owing to the low temperature of the filament much of the electric energy supplied is converted into invisible infra-red radiation. A search was made for substances of higher melting-point, which could be used at higher temperatures, when a greater proportion of the energy supplied is converted into visible radiation (Barton's "Heat," ch. 15), and both osmium and tantalum were tried without success. Finally, in 1909, Coolidge showed how to draw tungsten into very fine wires, and as its melting-point is about 3500° C. it is very suitable for the filaments of electric lamps. This doubled the efficiency, but the temperature of the filament could not be made higher, as it would evaporate and blacken the bulb. This was prevented by introducing an inert gas, argon, at about atmospheric pressure, so that the filament could be worked at a still higher temperature to increase the efficiency. This increase would have been offset by a loss in heat

due to conduction of heat through the argon. This was reduced to a minimum by a discovery made by Langmuir, who showed that a very thin filament was surrounded by a cylinder of hot stationary gas of constant diameter independent of the diameter of the filament and that the conduction loss took place off the surface of this cylinder. So he saw that the radiation produced by the filament could be increased without increasing the loss by conduction by coiling the filament so that it was like a spiral spring; the diameter of the coil was almost equal to that of the cylinder of stationary gas and the coils of the filament were practically touching one another. Thus the radiating area of the filament is now almost equal to the surface area of the cylinder of gas, whereas with a straight filament it is equal to the much smaller surface area of the filament itself. So the coiling increases the amount of radiation produced in a given time without increasing the conduction losses and is therefore equivalent to reducing the conduction loss for a given rate of production of radiation. This is the coiled gas-filled electric lamp and its efficiency is about four times that of the carbon filament lamp. These lamps give a continuous spectrum, but their light is deficient in blue and violet compared with sunlight, as the temperature of the filament is well below that of the surface of the sun.

89. ELECTRIC DISCHARGE LAMPS

An entirely new lamp has made its appearance in the last twenty years, which is rapidly superseding the filament lamp for industrial lighting, street lighting, and floodlighting. It is called the Electric Discharge Lamp, and the light is produced by the passage of an electric current through a gas. There are two distinct types. The cold cathode type is an ordinary discharge tube containing a gas such as nitrogen or neon at a pressure of a few millimetres of mercury and requires a P.D. of some 5000 volts; the electrodes and the tube are cold when it is working. The light is produced by the positive column in the usual discharge in a gas at low pressure, and since the length of this column is almost the same as that of the tube, this type of lamp can be used for outlining buildings and for night advertising signs. The colour of the lamp depends on the gas in it; neon gives a red colour, argon blue, carbon dioxide almost white. The light is produced in the following way. There are always some ions in a gas due to cosmic rays and traces of radio-active elements; when the electric field is switched on, these ions begin to move and, if the pressure in the gas is low enough, they may acquire sufficient energy between successive encounters to ionise a neutral molecule of the gas; that is, they knock an electron out of the molecule. This electron and the remnant of the molecule, now a positive ion, begin to move under the action of the field and they produce further ions. In this way a copious supply of ions is produced. But the positive ions are continually re-capturing electrons and, when they do so, electrical energy is transformed

into light. But it is not white light, but separate colours characteristic of the gas in the tube. In fact, the spectrum of the gas in the tube is emitted. These electric discharge lamps are not much more efficient than the vacuum tungsten filament lamp, their chief advantage lying in the variety of design available.

The second type of lamp, called the hot cathode type, has only become available commercially since 1931 and consists of an electric current at a P.D. of about 200 to 400 volts through a gas at pressures up to 1 atmosphere. The tube gets much hotter than in the first type. There is the **sodium vapour lamp** and the **mercury vapour lamp**. The sodium vapour lamp (Fig. 157) consists of a tube fitted with two electrodes, which are tungsten spirals coated with barium oxide to produce a copious electron emission. It contains neon at a pressure of 10 mm. of mercury and some metallic sodium on the walls of the tube, which is mounted in an exhausted vessel, so as to reduce to a minimum the loss of heat from

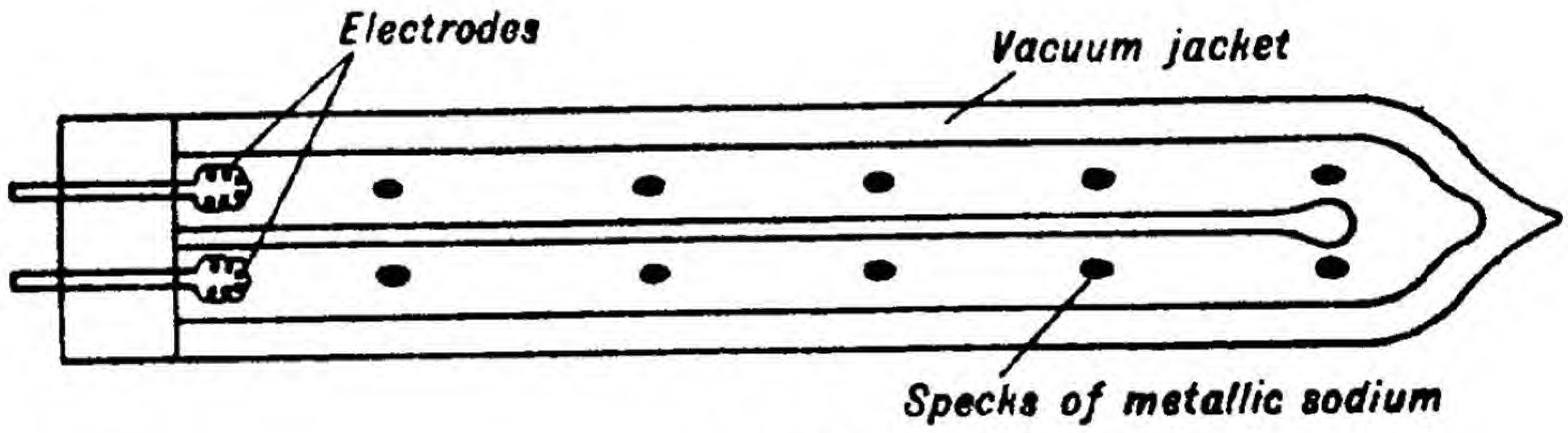


Fig. 157.

the hot inner tube. When the P.D. is switched on, the discharge passes through the neon, and the heat produced begins to vaporise some of the metallic sodium. In time, enough sodium vapour is produced to carry the whole of the current, sodium ions being produced in preference to neon ions, since the ionisation potential of sodium is less than that of neon. The sodium vapour is at a pressure of $\frac{1}{100}$ mm. of mercury when the lamp is working at its maximum output and the temperature of the tube is 275° C. Some of the electrical energy supplied to the lamp is used to keep the gas and tube at the correct working temperature and the rest is converted almost entirely into the well-known sodium D light, a small amount of infra-red light being produced as well. The efficiency of the lamp is about 50 lumens per watt, and it is the most efficient commercial lamp which has so far been produced ; but it is interesting to observe that only 10 per cent. of the electrical energy supplied is converted into visible radiation ! There is still room for improvement ! ! It is useless for lighting when colour discrimination is required, but the monochromatic character of its light is no handicap in street and industrial lighting.

The mercury vapour lamp works on the same principle, the discharge being started in argon at 10 mm. pressure and the tube containing a carefully measured mass of mercury so designed as to produce a pressure of 1 atmosphere when it is all vapourised at the working temperature of

the tube, 600° C. The visible radiation produced is in the yellow, green, and blue, and the lamp also emits a good deal of ultra-violet light. Its efficiency is about 40 lumens per watt, and it is suitable for street and industrial lighting. As it is deficient in red, while the filament lamp is deficient in blue and violet, a suitable combination of the two produces a close approximation to daylight; but this idea has not yet been used commercially to any extent. The mercury vapour lamp wastes some of its electrical energy in producing invisible ultra-violet radiation, and experiments are now being made to increase still further the efficiency of electric discharge lamps by coating the inside of the vacuum jacket with a **luminescent powder**, such as zinc orthosilicate, which will convert the ultra-violet radiation into light. The colour of the light emitted by the powder can be controlled by the trace of impurity which must be present to make the powder luminiscent and can be arranged to make the total light produced by the lamp more like daylight. The deficiency in red light of the mercury vapour lamp has been corrected to some extent by this means and the efficiency of the neon cold cathode lamp has been increased by 50 per cent. by the use of zinc orthosilicate. Here, then, is another example of the way in which scientific knowledge and scientific methods are being used to improve the material conditions under which mankind lives, and the reader who is interested to know more about these problems and their solution should consult the booklet on Electric Illumination published by His Majesty's Stationery Office, in which he will find a list of further books for reference.

TABLE 12

Type of Lamp.	Efficiency in Lumens per Watt.
Carbon filament.	3·5
Vacuum tungsten filament.	8
Gas-filled coiled filament (low power).	11–15
Arc lamps.	20
Cold cathode neon lamp.	15
Hot cathode mercury vapour lamp.	35–40
Hot cathode sodium vapour lamp.	40–55
Cold cathode neon lamp with luminescent powder.	22

90. THE PHOTOMETRY OF OPTICAL INSTRUMENTS

In discussing the design of optical instruments, we considered both their magnifying power and field of view, but we must now conclude this account of photometry by applying the principles we have established to the **brightness** of the images formed by optical instruments. There are two cases to be discussed, the one where the image is presented straight to the eye, as in the telescope and microscope, and the other where it is first cast on a screen, as in the camera and projection lantern. We must

rst establish two important relations concerning the emission of light rom bodies sending light out in all directions.

It is said that the brightness of a self-luminous body, such as a glowing piece of iron, or of a body which reflects diffusely, such as blotting-paper, is the same in whatever direction it is observed. Assuming this to be true, let us find how the rate of emission of light from such bodies must vary with direction. Let AC (Fig. 158) represent an elementary portion of such a surface of area dS whose brightness is B, so that the luminous intensity of the element of area in the direction of its normal is BdS. If the surface is observed from a direction making an angle θ with the normal to the surface, then the apparent area of the element AC in this direction is represented by AD and is equal to AC $\cos\theta$. Hence the luminous intensity of the surface in this direction must be $BdS\cos\theta$, if the brightness, which will be $\frac{BdS\cos\theta}{dS\cos\theta}$, is to remain constant. So, if I and I_0 are the luminous intensities of a surface in a direction θ with the normal and along the normal respectively,

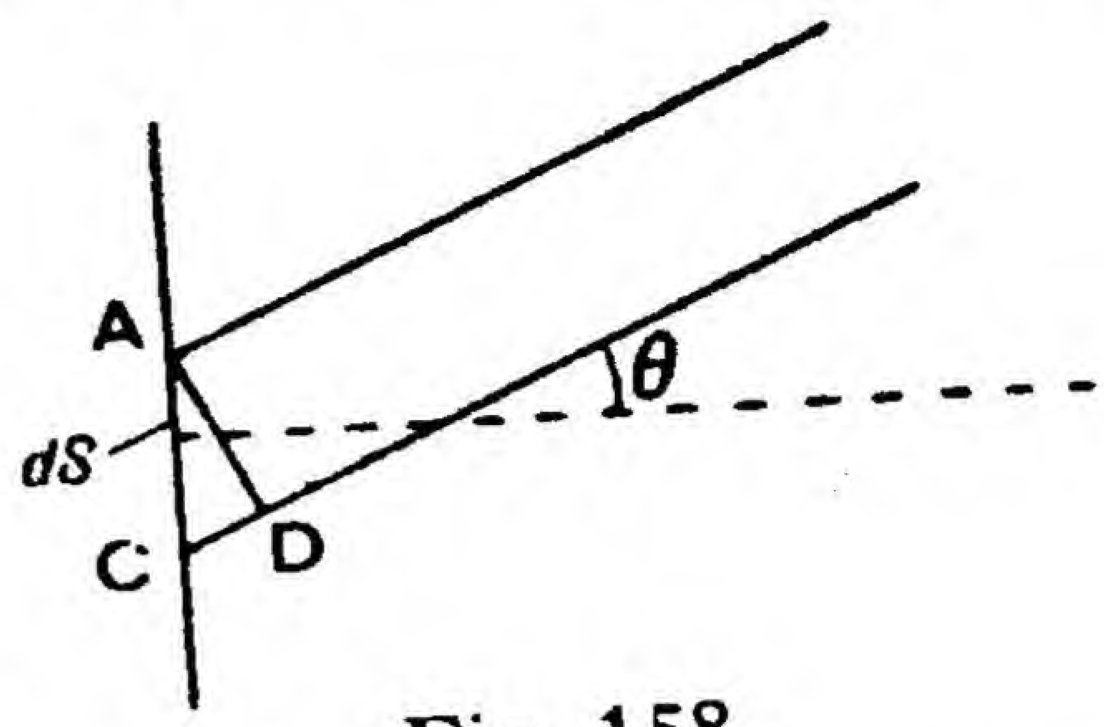


Fig. 158.

$$I = I_0 \cos\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

which is **Lambert's law of emission**. It must be emphasised that it is only true for surfaces which look equally bright from all directions. All surfaces are *assumed* to obey this law in what follows.

Let us now calculate the rate at which light is emitted by a surface of brightness B. Consider an element of area dS of the surface. The amount of light emitted in unit time lying between directions making angles θ and $\theta + d\theta$ with the normal is calculated by describing a sphere of unit radius about the element of area as centre. Then the portion of the surface of the sphere lying between the above directions is a zone of length $2\pi \sin\theta$ and width $d\theta$. Thus the light dF falling in unit time on this zone is given by

$$dF = BdS\cos\theta \,.\, 2\pi\sin\theta \,.\, d\theta$$

Hence the total amount of light emitted in unit time in a cone of semi-vertical angle θ is given by

$$F = \int_0^\theta 2\pi B \,.\, dS \,.\, \sin\theta \,.\, \cos\theta \,.\, d\theta$$

$$\therefore \quad F = \pi\, BdS \sin^2\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

If the surface is a plane emitting light into the whole of a hemisphere, then the total amount of light emitted in unit time per unit area is πB. The converse proposition is also important. If the illumination of a

perfectly diffuse surface is I, it acts as a source of light of brightness B, where

$$\pi B = I$$

This follows from the fact that each unit area of the surface must emit in unit time the same amount of light as falls on it in unit time.

91. THE ILLUMINATION OF THE IMAGE FORMED BY A SINGLE LENS

Let us consider an element of area dS_1 of an object of brightness B candles per sq. ft., of which an image of area dS_2 is cast on a screen by a lens (Fig. 159). The amount of light from the area dS_1 falling on the

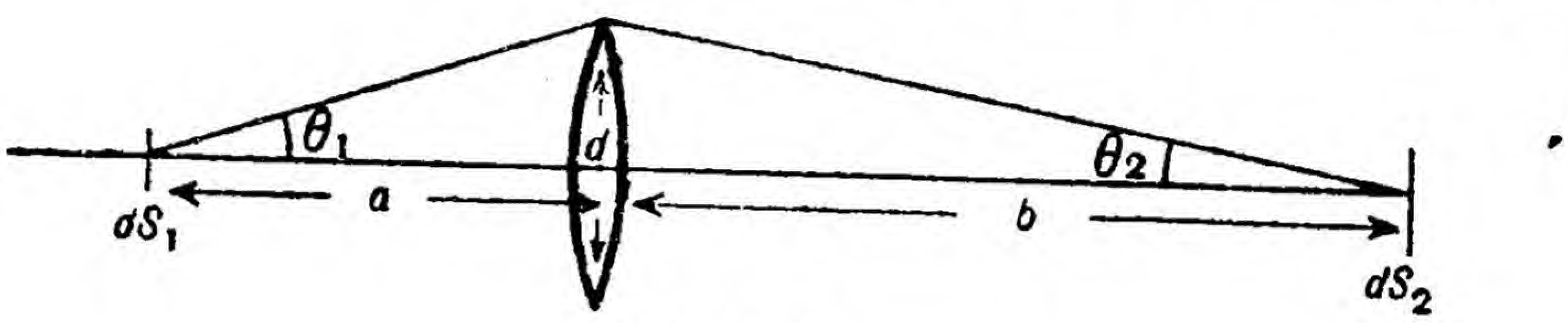


Fig. 159.

lens in unit time is $\pi B dS_1 \sin^2\theta_1$ from equation (50). If the lens transmits a fraction k of the light falling on it, k is called its transmission factor. Then the amount of light falling in unit time on the image is $k\pi B dS_1 \sin^2\theta_1$ and the illumination I of the image is given by

$$I = \frac{k\,\pi B dS_1 \sin^2\theta_1}{dS_2}$$

But $$\frac{dS_1}{dS_2} = \frac{a^2}{b^2}$$

and $$\sin^2\theta_1 = \frac{d^2}{4a^2} \text{ for small angles}$$

$$\therefore \quad I = \frac{k\pi B d^2}{4b^2} \quad \dots\dots\dots\dots \quad (51)$$

$$= \frac{kBA}{b^2} \quad \dots\dots\dots\dots \quad (52)$$

where A=area of the aperture of the lens, whose outline is assumed to be circular.

A very important proposition follows from this result. How bright an object appears when viewed by an observer depends on the illumination of the retinal image of the object formed by his eye. We shall call this brightness as judged by an observer's eye the **visual brightness** of an object to distinguish it from the brightness defined as candle-power per unit area. It is evident from equation (52) that the illumination of the retinal image of an object of brightness B is $\dfrac{tBA_p}{f^2}$, where t is the

transmission factor of the eye, A_p is the area of the pupil, and f is the distance from the eye lens to the retina, a constant quantity. So the illumination of the image is independent of the distance of the object from the eye, and depends only on the brightness of the object and the constants of the eye. Thus the visual brightness of an object of finite size is the same at all distances. The truth of this important result may be verified in another way. If an object is moved r times further from the eye, r^2 times less light enters the eye from it in each second; but the area of the retinal image is r^2 times smaller, so the illumination of the retinal image and the visual brightness of the object remain the same. This does not apply to point objects or objects which are so small that their image excites only one cone of the retina, because the removal of the object to r times its previous distance causes r^2 times less light to fall on the cone in unit time and so causes the object to appear r^2 times fainter than before. So we see that the brightness of an object of finite size—its candle-power per unit area—is a true measure of its visual brightness, since the distance away of the object has no influence on the illumination of the retinal image. Objects of the same brightness will appear equally bright to the same observer, even if they are at widely differing distances. The reason why distant objects appear fainter than near ones in practice is because of atmospheric absorption.

We shall now derive an expression for the brightness of the image of an object formed by a lens, when the image is not cast on a screen *but is to be observed directly by the eye*. Referring to Fig. 159 again, we have that the amount of light entering the lens in unit time is $\pi B_1 dS_1 \sin^2\theta_1$, where B_1 is the brightness of the object. Thus the amount of light entering the image in unit time is $k\pi B_1 dS_1 \sin^2\theta_1$, where k is the transmission factor of the lens. As the image is formed in air, the light from it is not spread out through an angle π, but only through a cone of semi-vertical angle θ_2. If B_2 is the brightness of the image, the amount of light emitted by it in unit time is $\pi B_2 dS_2 \sin^2\theta_2$.

$$\therefore\ k\pi B_1 dS_1 \sin^2\theta_1 = \pi B_2 dS_2 \sin^2\theta_2$$

$$\therefore\ B_2 = kB_1 \frac{dS_1 \sin^2\theta_1}{dS_2 \sin^2\theta_2}$$

$$= kB_1 \frac{a^2 \sin^2\theta_1}{b^2 \sin^2\theta_2}$$

If we restrict ourselves to small angles,

$$a^2 \sin^2\theta_1 = \frac{d^2}{4} = b^2 \sin^2\theta_2.$$

$$\therefore\ B_2 = kB_1 \qquad \text{. . . . . (53)}$$

It follows from this equation that **the brightness of the image of an object formed by a lens can never be greater than that of the object**

and is usually less on account of the light absorbed by the lens. Therefore the visual brightness of an object of finite size cannot be increased by the use of a lens. It must be emphasised that equation (53) applies only to images formed in air and not cast on a screen. If the image is cast on a screen, then its brightness B_s is calculated in this way. An amount of light $k\pi B_1 dS_1 \sin^2\theta_1$, falls on the image in unit time and an amount $\pi B_s dS_2$ is emitted in the same time.

$$\therefore k\pi B_1 dS_1 \sin^2\theta_1 = \pi B_s dS_2$$

$$\therefore B_s = kB_1 \frac{dS_1}{dS_2} \sin^2\theta_1$$

$$\therefore B_s = kB_1 \frac{a^2 \sin^2\theta_1}{b^2}$$

$$\therefore B_s = \frac{kB_1 d^2}{4b^2} \quad \ldots \ldots \ldots \quad (54)$$

This is less than for the aerial image since $d^2 < 4b^2$. It is quite clear that the brightness of an image cast on a screen is less than that of an aerial image, because in the former case the light is emitted over a hemisphere and in the latter only in a cone of semi-vertical angle less than 90°. We shall now apply these results to the telescope, microscope, and camera.

92. THE TELESCOPE AND MICROSCOPE

If the brightness of the object is B, then that of the final image is kB, where k is the transmission factor of the whole telescope. So the image seen through the telescope will be fainter than that seen by the naked eye, since an actual telescope may contain three lens systems and about 4 per cent. of the light incident on the surface is lost by reflection and absorption at each refraction. So altogether something like 25 per cent. to 30 per cent. of the light is lost in transmission through the telescope. But the visual brightness of the final image is only proportional to kB provided that the pupil of the eye is fully illuminated, that is, provided that the diameter of the eye-ring (Art. 65) is equal to or greater than that of the pupil. It is quite clear that increasing the diameter of the eye-ring beyond that of the pupil does not affect the visual brightness of the final image, because such an increase merely means that only a portion of the aperture of the objective is being used. But the brightness (and so visual brightness) of an image is independent of the diameter of the lens forming it, and so this decrease in the effective diameter of the objective does not affect the visual brightness of the object seen through the telescope. But, if the diameter of the eye-ring is less than that of the pupil, the illumination of the retinal image is obviously less than the maximum for an object of given brightness and so the visual brightness

is less than the maximum. If B is the brightness of the object, then the illumination of the retinal image formed by the naked eye is $\frac{tBA_p}{f^2}$ (Art. 91). The brightness of the final image formed by the telescope is kB, so that the illumination of the retinal image of this final image is $\frac{tkBA_r}{f^2}$, where A_r is the area of the eye-ring. This is due to the fact that the pupil of the eye is placed at the eye-ring, but light only passes through an area A_r of the pupil. Hence

$$\frac{\text{Illumination of Retinal Image with Telescope}}{\text{Illumination of Retinal Image with Naked Eye}}$$

$$=\frac{tkBA_r}{f^2}\cdot\frac{f^2}{tBA_p}=k\frac{A_r}{A_p}=\frac{kA_o}{M^2A_p}$$

where A_o=area of the objective and M=magnifying power of the telescope. Therefore the visual brightness of an object seen through a telescope is proportional to the area of the objective and inversely proportional to the square of the magnifying power, if the pupil is not fully illuminated.

The visual brightness of an object seen through a microscope is calculated in a similar way. If the brightness of the object is B, the amount of light falling in unit time from an area dS of the object on the pupil of the eye is $\frac{BdSA_p}{a^2}$, where a is the distance of the object from the eye. So the illumination of the retinal image of area dS_1 is $\frac{BdSA_p}{a^2dS_1}$. If the objective of the microscope subtends an angle 2θ at the object, then the amount of light from an area dS of the object falling on the objective in unit time is $\pi BdS\sin^2\theta$. All this light enters the pupil to form a retinal image of area M^2dS_1, where M is the magnifying power of the microscope. The illumination of this image is

$$\frac{\pi BdS\sin^2\theta}{M^2dS_1}$$

Hence $$\frac{\text{Illumination of Retinal Image with Microscope}}{\text{Illumination of Retinal Image with Naked Eye}}$$

$$=\frac{\pi a^2\sin^2\theta}{A_pM^2}$$

Thus the visual brightness of an image seen through a microscope is proportional to $\sin^2\theta$ and inversely proportional to the square of the magnifying power. If the objective is immersed in oil of refractive index n, it can be shown that the visual brightness is proportional to $(n\sin\theta)^2$, or the (Numerical Aperture)^2. This is another reason for having short-focu

objectives in a microscope, because the numerical aperture can be made larger with a short-focus lens than with a long-focus lens corrected to to the same extent for the various aberrations.

93. THE CAMERA

Many readers will have their own cameras and may be interested to know how the time of exposure varies with the aperture of the lens. If the diameter of the stop in front of the lens is $\frac{f}{r}$, where f is the focal length of the lens, r is called the stop number. Common values of r are 16, 11, 8, 6·3, 4·5. Assuming that the time of exposure needed is inversely proportional to the illumination of the plate, it is possible to find a relation between the stop number and time of exposure. If the camera is pointed at an object of brightness B at a distance a, the illumination I of the plate at a distance b from the lens is from equation (51)

$$I=\frac{k\pi Bd^2}{4b^2}$$

where d is the diameter of the stop in front of the lens. If the object is at a large distance, as is frequently the case, $b=f$ and

$$I=\frac{k\pi Bd^2}{4f^2}$$

If

$$d=\frac{f}{r}$$

$$I=\frac{k\pi B}{4r^2} \qquad (55)$$

Thus the illumination on the plate is inversely proportional to the square of the stop number and the exposure time is directly proportional to the square of the stop number. The other variable factor controlling the time of exposure is the brightness of the object to be photographed, and this is measured by an exposure meter.

It is interesting to calculate the illumination of the plate in one or two cases. If an object whose illumination is 100π foot-candles is being photographed, its brightness is 100 candles per sq. ft. from equation (53). So the illumination of the plate at a stop number of 11, assuming $k=1$, is given by

$$I=\frac{\pi\times 100}{4\times 11^2}=0{\cdot}65 \text{ foot-candles}$$

while, at a stop number of 4·3,

$$I=\frac{\pi\times 100}{4\times 4{\cdot}3^2}=4{\cdot}25 \text{ foot-candles}$$

which would reduce the time of exposure to one-seventh of its value at a

stop number of 11. Precisely the same type of calculation can be used to find the illumination on the screen of a projection lantern and some examples will be found at the end of this chapter.

The topics which have been discussed in this chapter illustrate one of the ways in which new knowledge is won. It is more than a hundred years since a method of comparing luminous intensities and a standard source of light were introduced, but interest in photometry has been revived in the last twenty years largely by the urgent social and industrial necessity of improving the lighting of public places. This has led to an improvement in the constancy and reproducibility of standard sources, to better methods of measuring luminous intensities and illuminations. Here great use is being made of the photo-electric effect, a phenomenon which was discovered and investigated in the first place solely in the pursuit of knowledge. These improvements in technique have made it possible to define scientifically the standards of illumination required and to devise means of realising them in practice. So the demands of society and industry have stimulated scientific activity and made use of scientific knowledge for the improvement of the working and living conditions of all classes of the community.

EXAMPLES ON CHAPTER IX

1. How would you compare the intensities of two sources of light? What difficulties arise in practice? (*Oxford Schol.*)

2. Describe the Lummer-Brodhun photometer, and explain how you would use it to compare the candle-powers of two sources of light.

Two lamps of 50 candle-power and 9 candle-power are placed at distances 50 cm. and 30 cm. respectively from a photometer and on the same side of the instrument. At what distance from the photometer must a third lamp of 27 candle-power be placed in order that both sides of the photometer may be equally illuminated? The lamps may be assumed to be on the normal to the photometer. At what distance would the third lamp have to be placed if the line joining it to the instrument were inclined at an angle of 60° to the normal? (*O. and C.*)

3. Describe, with the necessary theory, an accurate method whereby the illuminating powers of two sources of light may be compared. Suggest a form of apparatus suitable for comparing the reflecting powers of two plane mirrors. (*O. and C.*)

4. Describe concisely the Lummer-Brodhun photometer and some form of flicker photometer, stating the particular advantages in each case. (*London Inter.*)

5. Calculate the illumination on the ground midway between two lamp-posts 100 yds. apart and 15 ft. high, if the lamps are each of 400 candle-power, defining the unit in which illumination is measured. How would you check your result by observation? (*London Inter.*)

6. Mention the chief difficulties encountered in comparing the illuminating powers of two sources of light, and describe the apparatus and procedure you would use in order to obtain a satisfactory value of the ratio.

The table below gives the percentage of light which is transmitted by glass plates of varying thicknesses. The relation between the intensity of the incident light (I_0) and that of the transmitted light (I) is believed to be of the form $I = I_0\, a^x$,

where x is the thickness of the plate in mm. Use a graphical method to check this law and calculate the value of the constant a.

Thickness of the plate in mm. (x) .	0	4·2	8·5	12·0	16·4
Percentage transmitted	100	82·2	67·6	58·2	47·9

(*N.U.J.B.*)

7. Define the terms lumen and foot-candle, and show the connection between them.

Describe how you would compare the illuminating powers of two electric lamps. What special difficulty would arise if one lamp had a carbon and the other a metal filament ?

(*N.U.J.B.*)

8. Describe an experiment to compare the candle-powers of two lamps.

An intensity of illumination of 6 foot-candles is desired at a distance of 4 ft. from a lamp. What must be the candle-power of the lamp ?

If the most powerful lamp available was of half the required power, how would you place it in order to obtain the required illumination of 6 foot-candles ?

(*N.U.J.B.*)

9. How may the mean horizontal candle-power of an electric lamp be determined ? Give some account of the methods used to obtain an efficient distribution of light in three of the following : (a) searchlights, (b) street lamps, (c) hall lamps, and (d) ships' navigation lights.

(*London.*)

10. At what height should a light be placed in a room which has its walls, ceiling, and floor blackened, so that the intensity of illumination shall be a maximum at a point 5 cm. away from the point where the perpendicular from the source to the floor cuts the floor ? What would be the angle of incidence at that point ?

11. It is required to light a classroom 30 ft. × 20 ft. × 15 ft. to give a uniform illumination of 8 foot-candles on the desks. Taking the utilisation factor as 0·4 and the depreciation factor as 1·43, find how many 100-watt lamps of efficiency 12 lumens per watt will be needed.

12. Discuss the development of electric sources of artificial light.

13. State and prove Lambert's law on the variation of emission with direction.

The brightness of the sun is 165,000 candles per sq. cm. Calculate the illumination at the earth's surface at noon in lumens per sq. metre, given that the radius of the sun is 433,000 miles, the distance from the earth to the sun is 93,000,000 miles, and 10 per cent. of the light is absorbed by the atmosphere.

14. The illumination of the earth at noon is 9000 lumens per sq. ft. Calculate the brightness of a strip of ground covered with a white paint, which is a perfect diffuse reflector, either from first principles or proving any formula you use.

If you were given a 5000 candle-power floodlight, how high from the ground would you place it to produce the same illumination as the sun produces at noon ?

15. If the brightness of the sun is 154,000,000 candles per sq. ft., find the average illumination in lumens per sq. ft. over the half of the moon facing the sun assuming it to be 93,000,000 miles from the sun. Hence calculate the average brightness of the moon, assuming that if reflects perfectly diffusely 10 per cent. of the light falling on it.

16. How would you compare experimentally the intensities of two sources of illumination ?

The illumination of full sunlight at noon at the earth's surface is 100,000 lux. The lux is the illumination produced by a standard candle at 1 metre. If the moon were a perfectly white sphere, with a perfectly matt surface (i.e. giving no specular reflection), what would be the illumination at the earth's surface at full moon at midnight ? Explain your reasoning in full. (Apparent angular diameter of the moon = 15·5 minutes.)

(*Camb. Schol.*)

17. What is meant by the candle-power of a light source ?

A small mirror suitably curved is placed so that it projects to infinity an image of a small source of uniform brightness (i.e. candle-power per sq. cm. of projected area). Calculate the candle-power of the mirror considered as a source. (It will be necessary to take into account the divergence of the projected beam due to the finite size of the source.)

A motor car headlamp has a paraboloidal mirror with an aperture 10 in. in diameter, of focal length 2 in. At the focus is a spherical opal 24-watt bulb of uniform brightness $\frac{1}{2}$ in. in diameter. What is the candle-power of the headlamp? (The efficiency of the light may be taken as $\frac{1}{2}$ watt per candle. The mirror reflects specularly 80 per cent. of the incident light.) (*Camb. Schol.*)

18. Derive an expression for the brightness of an image cast on a screen by a lens in terms of the quantities you find necessary. Hence or otherwise show that the brightness of an object as seen by the eye is independent of its distance from the eye. Why do distant objects look fainter than near ones in practice?

19. Derive an expression for the brightness of an "aerial" image produced by a lens. Hence show that a lens can never increase the brightness of an object of finite size.

20. What do you understand by the intensity of a beam of light? How is it possible to compare the intensities of two sources of different colour? Explain carefully how you would make and calibrate a photographic exposure meter for use with electric light and sunlight. (*Oxford Schol.*)

21. Find an expression for the brightness as seen by the eye of the image of an object seen through a telescope in terms of the brightness of the object as seen by the naked eye. Use your expression to discuss the way in which a telescope is designed to produce images of the maximum brightness and show how a telescope causes point objects, such as stars, to look brighter than when seen by the naked eye.

22. Deduce an expression for the brightness as seen by the eye of an object seen through a microscope. Use your expression to show why a microscopic objective is a converging lens of *very small focal length*.

23. Prove that the illumination on the plate of a camera is inversely proportional to the square of the stop number. A camera is used to photograph a scene of brightness 400 candles per sq. ft. Find the illumination on the plate if the lens is stopped down to $f/16$, and the lens absorbs 20 per cent. of the light falling on it. If a cloud obscures the sun and it is found that the lens aperture must be opened up to $f/4{\cdot}3$ to use the same exposure as before, what is the brightness of the scene now?

24. What do you understand by the brightness of an object and of an image? On what factors does the brightness of an image depend?

How does the exposure required to take a photograph depend upon the diameter and focal length of the lens used? (*Oxford Schol.*)

25. A 500-watt lamp of efficiency 15 lumens per watt is used for projecting lantern slides. If the lamp is so far from the condenser that 1/10 of the light it emits falls on the condenser, and 10 per cent. of this is transmitted to the screen where the picture is 5 ft. × 5 ft., find the illumination on the screen. How would you verify your result experimentally?

26. Give an account of the optical system of a lantern suitable for the projection of lantern slides. Illustrate your answer with a diagram (or diagrams) showing the path of a pencil of rays (a) from a point on the source to the screen, (b) from a non-axial point on the slide to the screen. Why is it necessary to achromatise the projection lens and not the condenser?

The luminous flux incident on the condenser of a projector is 12,000 lumens, and the average illumination of the screen, 17 ft. square, is 5 foot-candles. Determine the fraction of the incident light transmitted by the optical system. (*N.U.J.B.*)

Chapter X

THEORIES OF LIGHT

94. THE FACTS

So far we have considered the theorems which may be deduced from the axioms formulated for rays of light and the optical instruments based on them and have not concerned ourselves with the nature of light. But there is always a body of men of a more reflective nature, who are not specially interested in instruments and machines but in the principles lying behind and governing natural processes. Such men naturally seek an answer to the question : What is light ? The answers to this question and the experimental evidence bearing on it are called **Physical Optics** to distinguish it from the preceding Geometrical Optics, but it is not strictly correct to separate them into watertight compartments. For we shall see that the knowledge of the nature of light has an important bearing on the design of optical instruments ; so that the debt which pure knowledge owes to industry in that industry stimulated it to erect the edifice of Geometrical Optics is repaid by the improvement in optical instruments due to the new knowledge of the nature of light (Art. 143).

We are going to discuss theories of light; that is, answers to the question: What is light ? It is well to begin by reminding ourselves of the purpose of a scientific theory and the lines along which it should be developed. The aim of a theory is to explain the experimental facts, to show how they are related to one another and to bring out their significance in a consistent scheme. An admirable example of this function of a theory is to be found in the Weber theory of magnetism, which explains quite naturally the impossibility of producing an isolated pole by supposing that magnetic substances consist of elementary magnets, an elementary magnet being the smallest magnet which can exist. In an ordinary piece of iron the axes of the magnets are distributed at random so that their resultant magnetic moment is zero, but when the iron is stroked with the N pole of a magnet all the south poles of the elementary magnets are attracted by it and tend to face the end of the iron at which the N pole leaves it. Hence the iron is turned into a magnet with a S pole at the end at which the N pole leaves the iron. Saturation occurs when the alignment is as perfect as can be. The simple facts of magnetic induction follow quite naturally and the whole theory gives a satisfying picture or working thought model of the process of magnetisation as

the re-arrangement of something which is already there, not the communication of a magnetic fluid to the iron. When a few simple experimental facts have been established and arranged in this way, they show how true it is to say that science is just ordered knowledge. But one other important thing is expected of a scientific theory: it must lead to a search for new facts; it must suggest new and fruitful lines of investigation. It is only in this way that science is raised above the level of mere trial and error, where progress is so painfully slow. A recent example of this is seen in the way in which the development of Wave Mechanics has stimulated experiments in the artificial disintegration of the atom. It is known that an atom consists of a minute positive nucleus, some 10^{-13} cm. in diameter, surrounded by a distribution of negative charge of sufficient magnitude to make the atom electrically neutral. The nucleus is the controlling factor in the atom and its mass and electric charge determine the atomic weight and atomic number of the atom respectively. In order to effect permanent disintegration of the atom, it is this nucleus which must be changed. Such a change is always going on before our eyes in the case of radio-active disintegration, which is accompanied in some cases by the emission of electrons and in others by the emission of α-particles, which are helium nuclei, whose speeds are such that it would require a potential of the order of a million volts to bring them to rest. Conversely it would require potentials of the same order to give α-particles sufficient speed to penetrate the nucleus against the repulsive force of its positive charge and so to disintegrate it. Therefore the late Lord Rutherford tried his first experiments on artificial disintegration with these α-particles and actually succeeded, but the chance of a direct hit with the nucleus is so small that very few atoms were disintegrated. It would be possible to produce more disintegration if artificially produced high-speed particles were available, since they can be generated in far larger numbers than α-particles. But this seemed quite out of the question, since it was not possible to produce the necessary high voltages artificially. But when the Wave Mechanics was put forward to explain the phenomena which occur inside the atom, one of its consequences was that there was a small but finite probability of particles of smaller energy than the maximum penetrating the nucleus; hence it would be possible to disintegrate a nucleus with voltages as low as 250,000 if only sufficient particles were available. This prediction of the Wave Mechanics was actually acted on by Cockcroft and Walton at Cambridge; they built a plant to speed protons through some 250,000 volts and on bombarding lithium with them they produced helium atoms. This was the first case of a complete artificial disintegration, in which both the projectile and the disintegration itself were produced in the laboratory, and it would never have been attempted but for the prediction of the Wave Mechanics. In seeking new knowledge, we are like travellers in search of treasure in an uncharted land; we need a signpost to lead

us to our goal, and the function of a theory is quite as much to act as that signpost as to explain and correlate already known facts.

There is one further thing about a successful theory. It must contain the minimum number of simple postulates and the remaining facts must follow by strict logical deduction. It must not be necessary to make a fresh assumption as each new fact turns up, firstly because such a state of affairs is intellectually unsatisfying, and secondly because such a theory would be quite unable to predict new facts. This is because it is not until the fact is discovered that we know what kind of an assumption has to be made to make the theory fit it. Assumptions made to fit new facts as they turn up are called *ad hoc* assumptions, and as soon as a theory has to resort to many of them, its end is in sight.

What are the simple fundamental facts about light which have to be explained by any theory as to the nature of light? We shall summarise them :

(*a*) Light is a form of energy. This is shown by the fact that, if a beam of light is spread out into a spectrum by a prism, the various colours cause a rise in temperature in a sensitive thermometer (Barton's Heat, Ch. 15), which shows that the light is converted into heat when it falls on any body which can absorb it. As heat is a form of energy, then so is light.

(*b*) Light travels through empty space.

(*c*) Light travels in straight lines. We have seen (Art. 2) that the evidence for this is not decisive, but there can be only a small departure from it.

(*d*) The laws of reflection (Art. 3).

(*e*) The laws of refraction (Art. 3).

(*f*) Simultaneous reflection and refraction. This refers to the fact that when a beam of light strikes a natural mirror, such as the surface of a pool of water, some is reflected while the rest is refracted.

(*g*) Light travels with a velocity of 3×10^{10} cm. per second and its velocity is independent of colour and the temperature of the source. The evidence for this statement is given in the succeeding chapter.

(*h*) White light consists of different kinds of light whose refractive indices in a given medium vary continuously between two given limits. To each given refractive index there corresponds a definite colour.

These were the main facts which were known and well established about the middle of the seventeenth century, when attempts were made to answer the question, What is light? It is true that one or two other facts were known, but they do not seem quite so fundamental and were not so well established as the ones given above, so we shall reserve discussion of them until a little later. The first two facts make it clear that there are only two possible types of theory, for there are only two possible ways of transmitting energy through empty space. One is by a stream of moving material particles, when the energy is transmitted

as the kinetic energy of the particles, and the other is by means of waves, when the energy is transmitted as the energy of the waves without any motion of the medium transmitting them. For example, if we wish to destroy or damage a ship at a distance, we may either fire a shell at it or we may produce a wave of large amplitude in the sea near to it. In the first case, it is the kinetic energy of the shell which damages the ship, and in the second it is the energy of the wave. And so two theories of light arose: one called the **corpuscular theory**, which states that light is a set of minute material corpuscles emitted by a luminous body which produce the sensation of sight on striking the eye; the other, called the **wave theory**, according to which a source of light sets up spherical waves in a medium called the ether which pervades the whole of the universe, in much the same way that a stone sets up circular ripples when it falls into a pond. These waves produce the sensation of sight when they enter the eye. We shall now discuss the explanation of the above fundamental facts about light by each theory.

95. THE CORPUSCULAR THEORY

The energy of a beam of light is the kinetic energy of the material corpuscles and their emission causes a loss in mass of the source; but this may well be too small to be detected, since their high velocity would produce an appreciable amount of energy even if the mass of each corpuscle were very small. It is quite clear that such corpuscles could travel through empty space. Indeed there is no difficulty in explaining this property of light, but it is not so easy to see how the corpuscles can travel through solids and liquids. Presumably they are so small as to be able to pass through the interstices of matter. The rectilinear propagation of light follows from Newton's First Law of Motion, for the force of gravity is too small to produce any noticeable deviation from straight line propagation in the length of a room in a particle going at 3×10^{10} cm. per second.

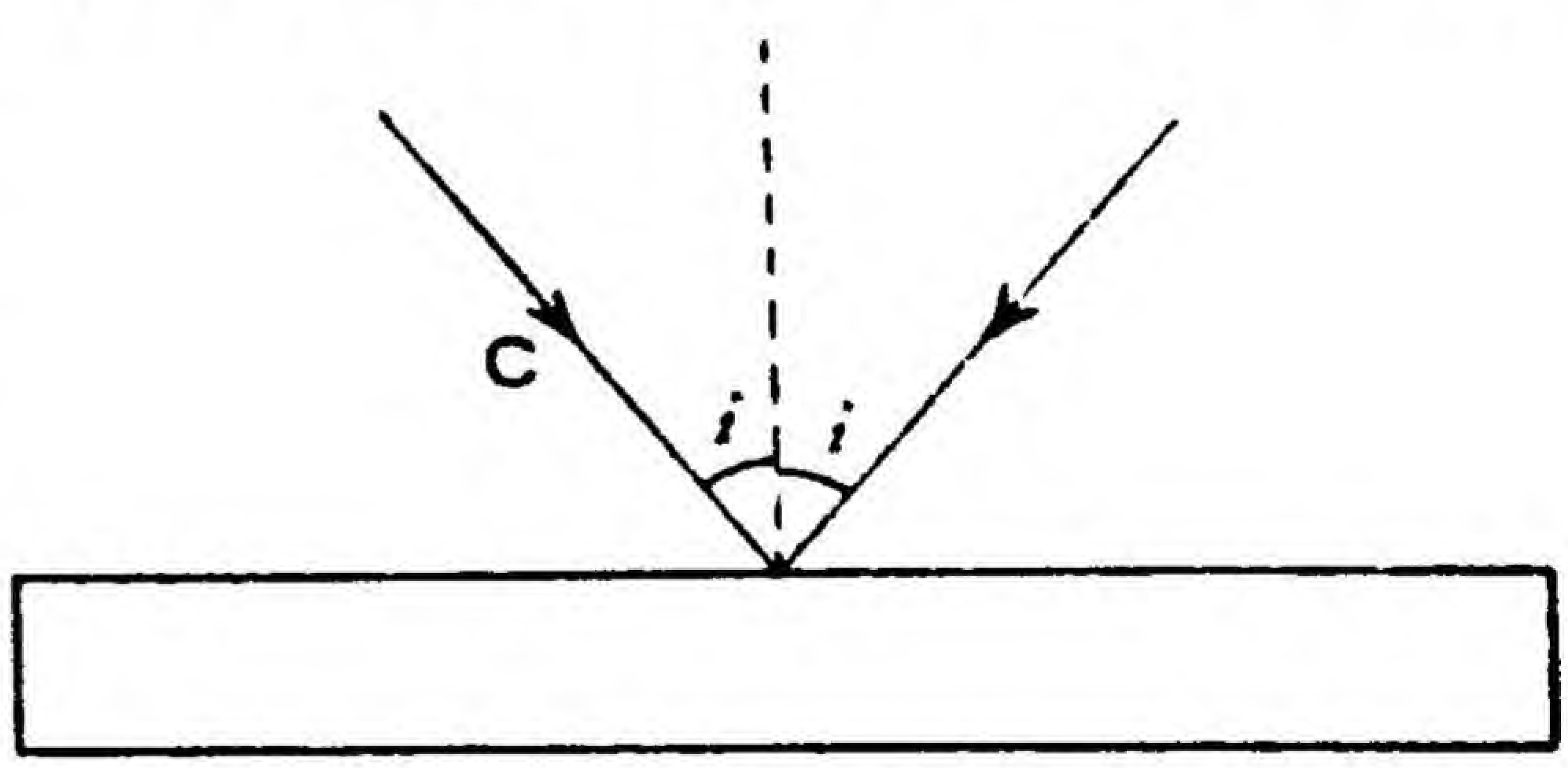


Fig. 160.

The laws of reflection follow at once, if the corpuscle suffers a perfectly elastic impact with the mirror. If the corpuscle is incident at an angle i at a velocity c (Fig. 160), the tangential component $c \sin i$ is unchanged, while the normal component $c \cos i$ is exactly reversed in magnitude and direction and so the two components add to a velocity c, making an angle i with the normal on the opposite side and in the same plane as the incident ray and the normal. So both the laws of reflection follow from this picture of an elastic impact between the corpuscle and the mirror. But we must admit that we have not yet explained *why* such an impact should occur and *why* it should be perfectly elastic.

Refraction is accounted for in the following way. Let us suppose that a corpuscle moving with velocity c_w in a medium such as water is incident at an angle i_2 at the surface and is refracted into air so that the angle of inclination in air is i_1 (Fig. 161) and the velocity is c. The ray bends away from the normal when it enters the air, because the normal component of the velocity is decreased by the attraction of the molecules of the water for the corpuscle, while the tangential component is unaltered, since there is no resultant attraction on the corpuscle due to the water in a direction parallel to its surface. We have at once that

$$c_w \sin i_2 = c \sin i_1$$

since the tangential component of the velocity is unaltered.

$$\therefore \frac{\sin i_1}{\sin i_2} = \frac{c}{c_w}$$

a constant for the two media, which is Snell's law. Thus the corpuscular theory gives a satisfactory quantitative explanation of refraction. The attraction is molecular rather than gravitational, since it must start and finish very close to the surface to account for the sudden bending and the gravitational law of inverse squares would permit of gradual bending. So we must postulate a molecular attraction, since it ceases at very small distances, as the facts of surface tension show us. We notice that this explanation of refraction requires *the velocity of light in water to be greater than that in air.* This prediction is a good feature of the theory, for it is suggesting fresh experiments and is lending itself to further test. We see that the theory gives a suggestive picture of total internal reflection, which occurs at such oblique incidence that the normal component of the velocity is not big enough to enable the corpuscle to overcome the molecular attraction.

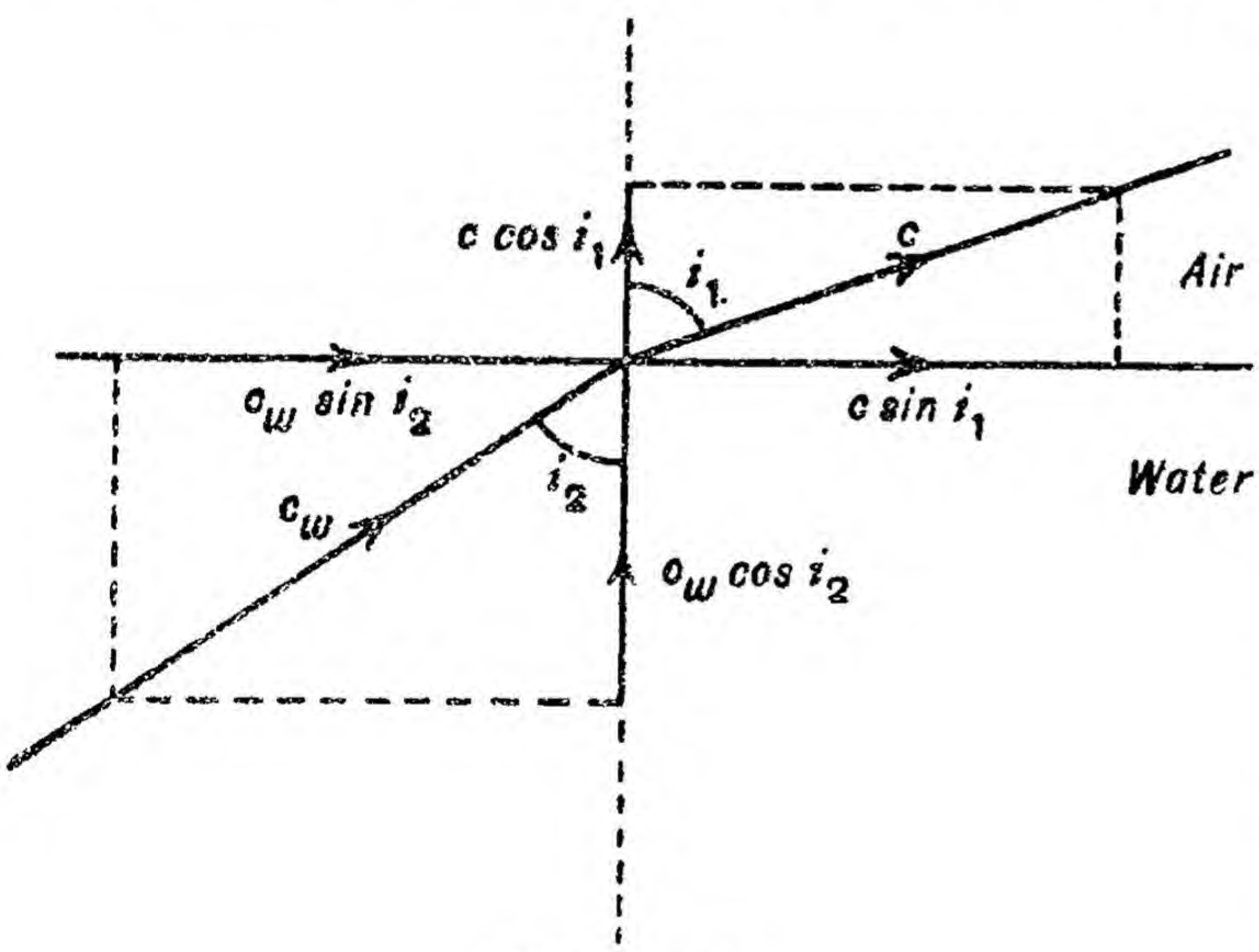


Fig. 161.

So far it has been a triumph for the corpuscular theory, but what about simultaneous reflection and refraction? Why are some of the corpuscles reflected and some refracted? The fact is that we have some explanation of refraction, but we have none of reflection; we have no repulsive force to account for the elastic impact. This is not an insuperable difficulty, for in the molecular theory of matter the molecules attract one another at ordinary distances but must repel one another at very

close distances. We now know that this is due to the outer charge of negative electricity around the nuclei of the atoms, but it must have been a difficulty before the structure of the atom was discovered. A solution may be found to our present difficulty, but for the moment we must frankly recognise that the theory has failed to explain simultaneous reflection and refraction.

We should not expect the theory to predict the precise velocity of light, but it is natural to suppose that the higher the temperature of the source the faster the corpuscles would be emitted from it. Still greater force is lent to this argument when we realise that the higher the temperature, the greater the average kinetic energy of the molecules of the source. But the velocity is found to be the same whatever the temperature of the source and so the corpuscular theory has failed again. It offers a satisfactory explanation of the colours of the spectrum, which are due to different sized corpuscles.

96. THE WAVE THEORY

The wave theory asserts that a luminous body sets up waves in an all-pervading medium called the **ether** and that these waves produce the sensation of sight on striking the eye. In science we often proceed in the first place from the known to the unknown, although the final form of a theory is expressed in abstract terms. So we shall content ourselves for the present by discussing if the two known kinds of waves, water waves and sound waves, obey the same type of laws as light. If they do, it is reasonable to believe that light may be waves. In other words, for the moment, our explanation of a fact about light on the wave theory is just the existence of a corresponding fact in water waves and sound waves. Both of these waves carry energy, and so the first fact is at once explained. But both water waves and sound require a material medium, and so how can light travel through empty space ? It is all very well to fill the whole of space with a medium just for the purpose of transmitting light waves, but is this not an *ad hoc* assumption ? This is a serious point and will be discussed a little later. Again, how is rectilinear propagation to be explained, for both water waves and sound waves spread sideways into the space behind obstacles, and it is impossible to cast sharp shadows of either ? The large waves on the sea pass by a boulder and, although the wave is temporarily divided by the obstacle, in a yard or two the two parts of the wave have joined together again to form one continuous crest, which bears no trace of having passed by a boulder. The wave has spread sideways at each side of the obstacle and reformed. In the same way, we may get out of the sight of the preacher in a church by getting behind a pillar, but we cannot get away from the sound of his voice, because the sound waves spread sideways after passing the pillar and come to our ears ! The direction in which any small portion of the wave travels is called a ray and the rays are perpendicular to the waves ;

in each of the above two cases the rays bend round the corner of the obstacle. **This sideways spreading of the waves or bending of the rays round obstacles,** which is so characteristic of both water waves and sound waves, is called **diffraction.** What more striking difference could there be between the behaviour of light and sound in the above example, the one exhibiting diffraction, the other travelling in straight lines? And was it not this striking contrast between sound and light which made Newton think: "I know sound is waves, how then can light be waves if it is so different from sound in this matter of diffraction?" Newton felt that this matter of rectilinear propagation was the essential fact about light, the one fact which was the real key to its nature and it was this which made him reject the wave theory in favour of the corpuscular theory.

So far there seems to be little to be said for the wave theory, but it gives quite a natural explanation of reflection and refraction. It is possible to study water waves by producing them artificially in a ripple-tank, which is a rectangular trough with a glass bottom filled with water to a suitable depth. The waves are seen by casting an image of the water

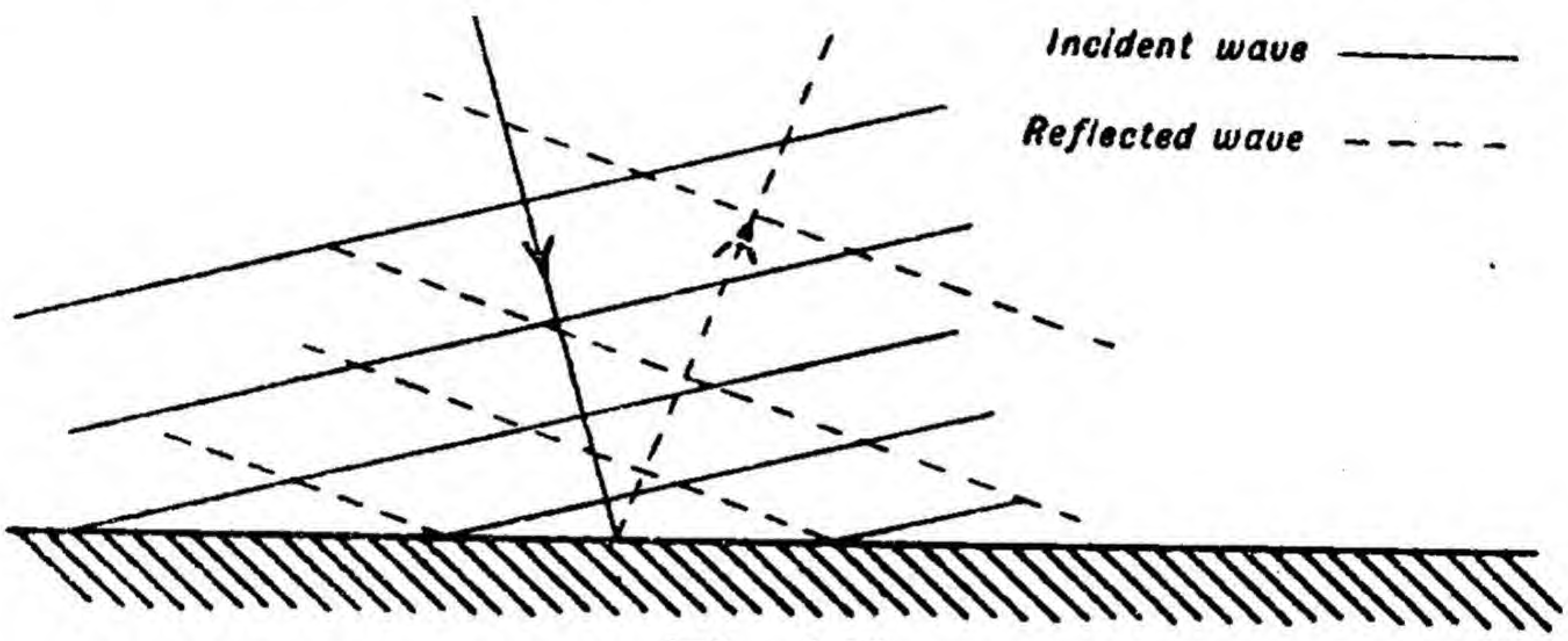


Fig. 162.

surface, which is suitably illuminated, on to a screen. If straight waves (Fig. 162) are produced by making a strip of metal about 3 in. long vibrate up and down in the water surface, they can be reflected from a straight piece of wood. It can be seen that the reflected waves make about the same angle with the mirror as the incident ones. The rays are perpendicular to the wave fronts and so this verifies roughly the law of reflection. We should notice that the wave theory has given no reason why light should be reflected; it just contents itself with saying, for the moment, that water waves can be reflected by suitable surfaces and, when they are, they obey the same laws as light. It is also known that a wave from a point source of sound is reflected by a plane surface so as to produce a point image as far behind the surface as the source is in front, which is precisely the same as is found with a point source of light and a plane mirror. Hence if light is waves, from what we know of water and sound waves, it would obey the observed laws of reflection.

Refraction is quite a well-known phenomenon in water waves and sound. If straight water waves are produced in a ripple tank and are incident obliquely on the line of separation between deep and shallow

water in the tank, the waves wheel round as shown in Fig. 163, due to the fact that they travel more slowly in shallow than in deep water. This wheeling of the waves is equivalent to the bending of the rays as they pass from deep to shallow water and is just the same as the bending of light rays when they pass from air to glass, for example. Precisely the same phenomenon occurs in sound. Sound travels better with the wind

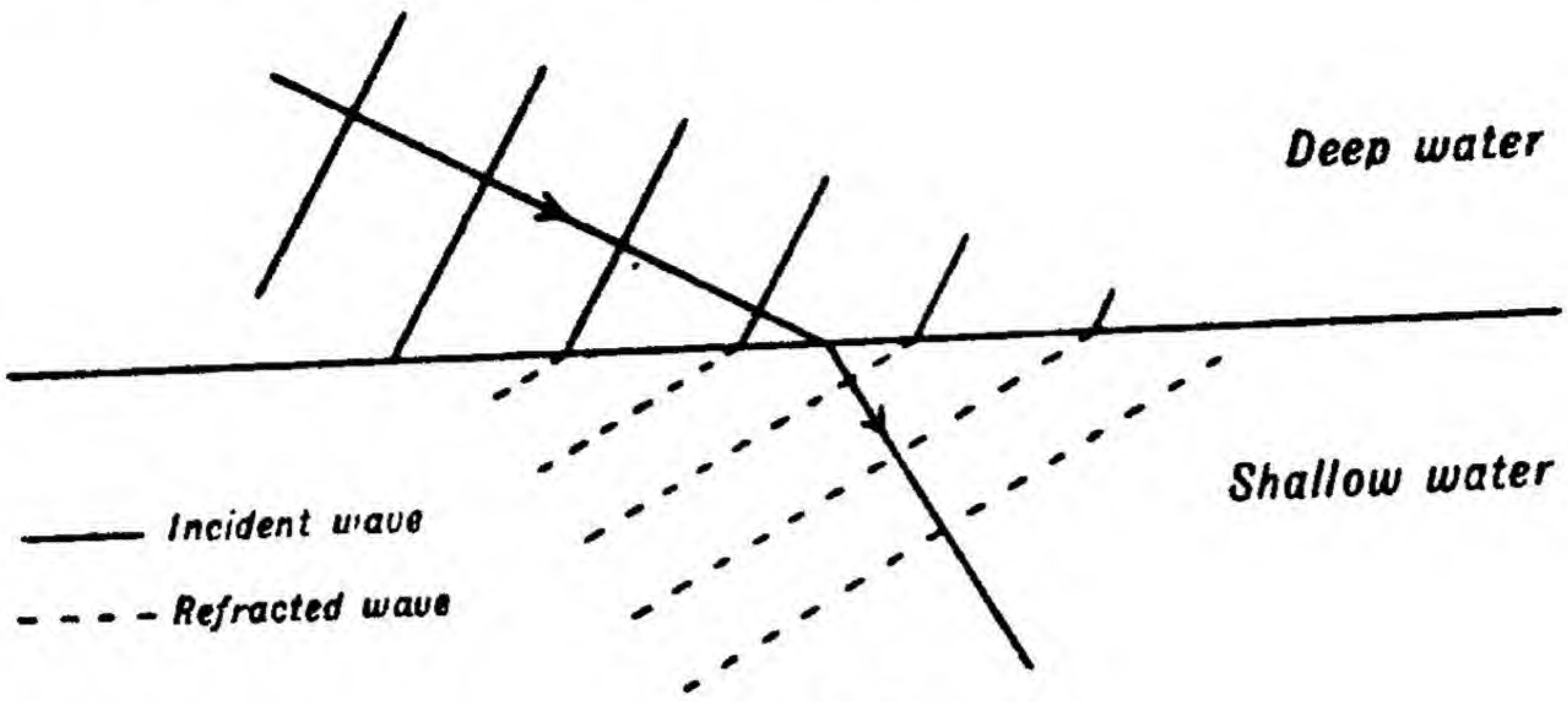


Fig. 163.

than against it, because the wind velocity is greater some distance above the ground than close to it. The velocity of the sound relative to the ground is the algebraic sum of the velocity of sound in still air and the wind velocity, and so the sound waves travel more rapidly some distance above the ground than close to it in the direction of the wind, while the opposite is the case against the wind. The result is that the waves wheel down towards the ground in the direction of the wind and wheel upwards from the ground against the wind. Therefore the sound rays curve down towards the ground with the wind so that the sound is abnormally loud, whereas they curve upwards away from the ground against the wind so that the sound is abnormally faint. **Refraction, then, is a wheeling of the waves (and a bending of the rays) due to one end travelling at a different velocity from the other since the properties of the medium vary along the length of the wave.** Precisely the same explanation can apply to the refraction of light, remembering that the waves are normal to the rays. The bending of the rays towards the normal when light passes from air to a dense medium such as water is due to the waves wheeling round so as to be more nearly parallel to the refracting surface (Fig. 164). *It is clear that this involves light travelling*

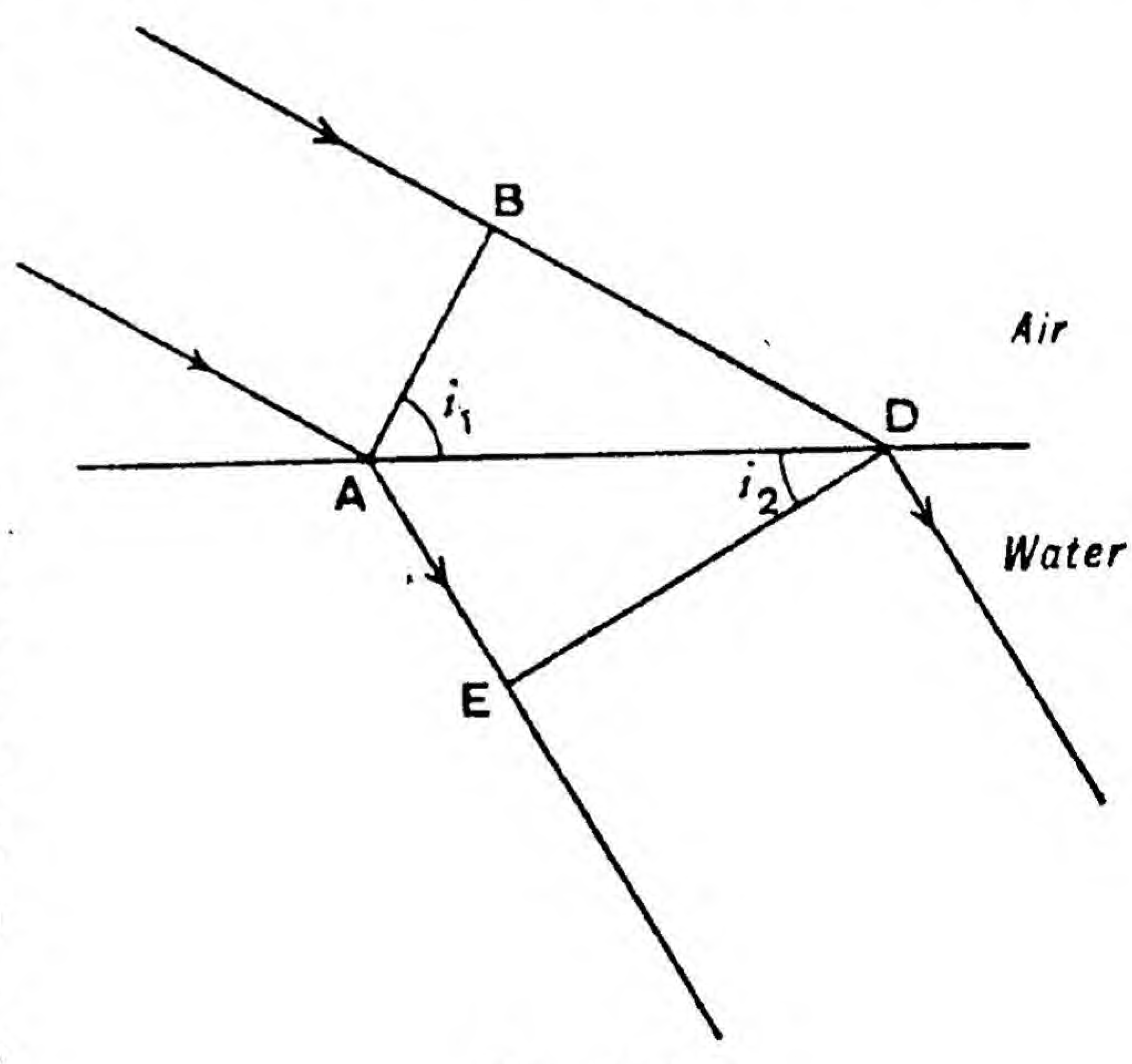


Fig. 164.

more slowly in water than in air, the precise opposite of the corpuscular theory. But like that theory, the wave theory leads to Snell's law. If the incident wave AB wheels round to become the refracted wave DE, then light travels a distance BD in air while it is going AE in water. If c and c_w are the velocities of light in air and water respectively, we have

$$\frac{c}{c_w}=\frac{\mathrm{BD}}{\mathrm{AE}}=\frac{\frac{\mathrm{BD}}{\mathrm{AD}}}{\frac{\mathrm{AE}}{\mathrm{AD}}}=\frac{\sin i_1}{\sin i_2}$$

$$\therefore \frac{\sin i_1}{\sin i_2}=\frac{c}{c_w}$$

a constant for two given media, which is Snell's law. The wave theory does not offer any explanation of why light should travel more slowly in water than in air, but merely asserts that this must be so to explain refraction. The clear-cut disagreement in the interpretation of refraction by the two theories suggests that attempts should be made to measure the velocity of light in air and water, so as to help to decide between the two theories.

Can the wave theory offer any explanation of simultaneous reflection and refraction? It can, in the sense that it is well known in sound and the reason for it is understood. If a sound wave travels along a cylindrical column of air, of which one end is closed and the other open, the wave will be reflected at the open end. In this way stationary waves are set up in the pipe and account for the well-known relation between its length and frequency and also for the set of overtones accompanying the fundamental. If the wave approaching the open end is a compression, it is reflected as a rarefaction, simply because the pressure must remain constant at the open end. In practice, however, this condition is not strictly fulfilled, and the node of pressure variation is a little above the end of the pipe and some of the sound is reflected back into the pipe and some is transmitted into the surrounding air. It is just as well that this is so from the point of view of the organist, since it would not be very entertaining to produce vibrations of the air inside the organ pipes if none of the vibrations was transmitted to the ears of the audience! Presumably there is a similar explanation of reflection in the case of light and so simultaneous reflection and refraction is quite possible.

As in the corpuscular theory, we should not expect to be able to account for the precise value of the velocity of light, but it is natural to expect the velocity to be independent of the temperature of the source, since the velocity of waves is settled entirely by the properties of the medium. Finally the wave theory accounts for the different colours into which white light can be split by a prism by asserting that each colour has a different frequency. This amounts to saying that the wave-lengths

of different colours in the same medium, such as air, are different. It is necessary to make this restriction, since the wave length of given colour varies as it goes from medium to medium, as is clear from Fig. 163.

97. FURTHER EVIDENCE

Before summing up the merits of the rival theories, we shall allude to three phenomena which were just discovered about the time that the theories were put forward, but which were not so well known or so definitely established as the facts whose explanation has been attempted above. The three phenomena are Newton's rings, Grimaldi's experiments on shadows, and Bartholinus' observation of double refraction. Commencing with Newton's rings, we shall not describe the observations precisely as he made them but in the form in which the experiment is usually arranged nowadays. If a beam of monochromatic light, such as sodium yellow light, is sent on to a long-focus convex lens resting on a plane glass plate, a set of alternate dark and bright rings is observed with the point of contact of the lens and plate as centre (Plate IV). The centre of the rings is a dark spot and the rings get closer together as we go outwards, their radii being proportional to the square root of the natural numbers. Simple geometrical considerations (Art. 129) prove that the thicknesses of the air film between the lens and plate, at which dark rings occur, are in the ratio of 0, 2, 4, 6, 8, and so on, that at which bright rings occur being in the ratio 1, 3, 5, 7, and so on. Hooke had obtained similar effects with thin plates and considered they were due to light reflected from both the front and back surfaces. He supported this contention by showing that no such effects were obtained if an opaque plate were used, thus suppressing the reflection from the back surface. But Newton took the view that all Hooke had shown was that reflection from the back surface was essential and his explanation of Newton's rings ignores the light reflected from the upper surface of the air film. Newton felt that this recurrence of darkness at regularly increasing thicknesses of the film suggested that light was of a periodic nature and a periodic nature suggests waves. How is this to be reconciled with a corpuscular theory? The picture of gravitation of those days also filled the whole of space with a medium called the ether, which was densest in space and least dense in the presence of matter. This accounted for gravity because matter was supposed to try to go to places of minimum ether density, that is, to places occupied by other pieces of matter. The reflection of a ray of light in water when it strikes the surface separating the water from the air (Fig. 161) occurs because the corpuscle is reflected when it encounters ether more dense than that in which it is travelling. When the corpuscles strike the upper face of the air film between the lens and plate used to form Newton's rings, this produces waves in the ether very much as a stone striking the surface of a pond produces ripples on the surface. These ether waves, which are regions of varying ether

density, travel faster than the corpuscles and throw a given corpuscle first into a **fit of transmission**, when a region of dense ether is around it, and then into a **fit of reflection**, when a region of rare ether is around it. Hence the corpuscle will travel a distance D in air between two successive fits of transmission. Hence the dark rings occur at thicknesses of the film equal to D, 2D, 3D, and so on. For consider the thickness D. If the corpuscle starts at the top of the film in a fit of transmission produced by striking the upper surface of the film, it goes through this surface and reaches the lower surface, which is a distance D below it. Hence it is again in a fit of transmission and so is transmitted through this surface too. Hence no light is reflected from that part of the film which is of thickness D, so that a dark ring of appropriate radius is formed. The same argument applies to the other dark rings, there being r fits of transmission between entering and leaving the film where the thickness is rD. The bright rings in between occur at thicknesses of D/2, 3D/2, 5D/2, and so on. Here any corpuscle penetrating the upper surface of the film due to being in a fit of transmission reaches the lower surface in a fit of reflection and so is returned to the upper surface again and reaches it in a fit of transmission and so emerges and enters the eye. Newton's rings do suggest very strongly that light has a periodic nature, and Newton was scientific enough to see this and to frame his explanation accordingly. But the corpuscle is introduced because it is so difficult to see how light can be waves in view of rectilinear propagation. He actually recorded that the distance between two fits of transmission is greater for red than for blue light and that the distance for yellow light is $\frac{1}{89,000}$ of an inch or $2{\cdot}86\times10^{-5}$ cm.

It is now possible to suggest some explanation of simultaneous reflection and refraction. For the corpuscles will usually be accompanied by the ether waves going at a greater speed than themselves. If the particle in air strikes the surface of a medium such as water when it is surrounded by ether of low density that ether may be even less dense than the ether in the water and so reflection will occur, since the corpuscle is attracted to the place of minimum ether density. If, on the other hand, it is surrounded by ether of large density, its density will be greater than that of the ether in the water and so transmission will occur.

Grimaldi was the first person to make experiments on shadows using small sources of light and performing actual measurements. His source was a tiny aperture in a blind through which sunlight passed into a dark room and he found that the actual shadow of a small obstacle was larger than the geometrical shadow. This must be a hard experiment to do, as it is difficult to tell exactly where the shadow ends, and using a motor car headlight with a line filament end-on and working at distances of 20 cm. or so between the source and obstacle and obstacle and screen, I have found that the shadow is smaller than the geometrical shadow.

But Grimaldi also noticed that there are three dark bands parallel to the edge of the shadow outside it, the bends getting *closer together* the further they were from the shadow. He also noticed *equally spaced* dark and bright bands inside the shadow. Newton repeated these experiments and confirmed the first two sets of observations, but failed to notice any bands inside the shadow. Are these observations to be regarded as facts? Is the evidence sound enough? And if so, do they indicate that light does not really travel quite in straight lines? If we accept the evidence, it is clear that light does not quite travel in straight lines, since the shadow is a little larger than the geometrical shadow. This can be explained by supposing that the region of less dense ether in the obstacle extends a little beyond the boundary of the obstacle itself, so that corpuscles just missing the obstacle would be reflected at the surface of this less dense ether, thus making the shadow rather larger than the geometrical shadow and explaining why the corpuscles did not travel quite in straight lines. But what is the explanation of the bands outside or inside the shadow?

We shall not describe the precise facts of double refraction discovered by Bartholinus, but some rather simpler ones discovered later of which the interpretation is the same. If a ray of light is sent normally on to a crystal of calcite, it is split into two rays by the refraction, one obeying the ordinary laws of refraction and the other behaving in a different way. If either of these rays is sent through a suitable instrument, such as a Nicol prism (Art. 154), it is found that, if the Nicol prism is rotated about the ray as axis, twice in a revolution the intensity of the emergent ray is zero and at two other intermediate positions it is a maximum. The ray is "one-sided," to use Newton's term. This last piece of evidence seems to favour the corpuscular theory, for it is known that sound waves are longitudinal and exhibit no such one-sidedness. And must not any waves in a medium filling the whole of space be the same? It is true that water waves are one-sided, but they occur only in two dimensions and cannot be used to explain the above fact about light. This **one-sidedness** of light or its **polarisation**, to use the modern term, is explained on the corpuscular theory by having oblong corpuscles and regarding the Nicol prism as a kind of slot, which only passes the corpuscles when their length is parallel to the slot.

98. SUMMING UP AND SUGGESTIONS FOR FURTHER EXPERIMENTS

We may now sum up the position of the two theories in the light of the evidence which we have so far obtained and see in what way we can hope to decide between them. The corpuscular theory explains successfully why light travels through empty space, and the facts of rectilinear propagation, reflection, refraction and colour. It has no explanation of the independence of the velocity of light of the temperature of the source,

and its explanation of Newton's rings and simultaneous reflection and refraction involves the introduction of waves and gives to light a dual nature. Yet again how can light be waves if it is one-sided? And this brings us back to rectilinear propagation, which also seems to rule out a wave motion.

And what of the wave theory? It is successful in explaining reflection, refraction, simultaneous reflection and refraction and colour. It leads quite naturally to the velocity of light being independent of the temperature of the source. But can we believe in its explanation of why light can travel through empty space? Can we believe in this medium which fills the whole of space and yet cannot be detected by our senses? Can we believe in the remarkable properties which it must have to explain the great velocity of light and the fact that the planets move through it without experiencing any resisting force? The velocity of longitudinal waves in a medium is $\sqrt{\dfrac{\text{Elasticity}}{\text{Density}}}$ and to obtain a value of 3×10^{10} cm. per sec. *the elasticity of the ether must be many times greater than that of steel and its density many times less than that of the best vacuum we can produce!* And its viscosity must be negligible to account for the fact that the planets show no signs of slowing down in their motion through it round the sun. Finally, when Michelson and Morley attempted to measure the velocity of the earth through the ether in 1888 they received the answer $\frac{0}{0}$, or anything! Can we believe in this elusive medium, whose properties are so bizarre and which is of such a retiring disposition that it runs away whenever we try to make measurements on it? This is certainly a difficulty in the way of accepting the wave theory. And then what of rectilinear propagation? Are the experiments of Grimaldi really evidence that rectilinear propagation is only an approximation to the truth and are the bright bands inside the shadow evidence of light rays bending round the corners of the obstacle? It may be that they are, but there is a big difference between these bands and the sweeping round the corners of obstacles, which occurs with water waves and sound waves.

It is clear that neither theory is wholly satisfactory and that each suggests further experiments, which may provide the additional evidence needed to decide between them. The present evidence is clearly unable to do this. Our next steps are quite clear. First we must learn to measure the velocity of light by laboratory methods so as to settle whether light travels more slowly in air than in water or the other way round. Secondly, we must investigate the properties of waves carefully and place them on a sound theoretical foundation to get an exact mathematical explanation of reflection and refraction. We may then be able to find some explanation of Newton's rings on the wave theory and we may also be led to set up further experiments to illustrate the undulatory character

of light. Thirdly, we must explain on the wave theory why rectilinear propagation is so nearly true, we must clear up Grimaldi's experiments, and *we must demonstrate without any doubt the diffraction of light*, if the wave theory is to survive. We shall deal with these problems in the succeeding chapters.

EXAMPLES ON CHAPTER X

1. State the fundamental facts about light and discuss their explanation on the corpuscular theory and the wave theory.

2. If the only facts known about light were rectilinear propagation and the law of reflection, would you prefer the wave theory or the corpuscular theory? Give reasons for your answer.

3. Compare and contrast the explanations of refraction on the corpuscular theory and wave theory. If experiment should decide against the wave theory, would you regard this fact as fatal to the wave theory? If not, what would you try to do in order to establish the wave theory again? (*London.*)

4. Write a short essay on the nature of light.

Chapter XI

THE VELOCITY OF LIGHT

99. GALILEO'S ATTEMPT

We have seen that it is important to be able to measure the velocity of light by a laboratory method in order to decide if light travels the more slowly in air or water, but there is also an intrinsic interest in the velocity of light itself. It is obvious that light travels very fast, but is its propagation infinite or does it travel with a finite velocity? Kepler and Descartes held the former view, but Francis Bacon and Galileo supported the latter, and Galileo was the first person to try to settle the question. He stationed two observers, A and B, a few miles apart, each provided with a lamp covered with a shutter. A uncovered his lamp at the same time starting a clock. As soon as B saw the light of A's lamp he uncovered his lamp and A stopped the clock as soon as he saw the light of B's lamp. The time recorded on the clock is the time taken for light to travel from A to B and back again together with the time taken by B to answer his lamp after seeing the light of the other lamp and the time taken by A to stop the clock after seeing the light of B's lamp. The total time taken by A and B to perform these operations is found by repeating the experiment with A and B close together so that the time taken by the light in going from A to B and back is negligible. It was found the total time recorded on the clock was the same whether A and B were close together or several miles apart. In other words, the time taken by light to go from A to B and back when they were several miles apart was zero! It is interesting to notice Galileo's conclusion. It was not that the velocity of light was infinite, but that it was too great to be measured by this method. It is true that the method is somewhat crude judged by present-day standards, but it was probably the best that could have been devised with the very simple apparatus available in Galileo's time, and we shall see that it embodied the essential principles of the succeeding methods.

100. RÖMER'S ASTRONOMICAL METHOD

We have already seen (Art. 66) that Galileo was one of the first persons in Europe to make a telescope, and one of the first things he did with it was to point it at the planet Jupiter, and to his delight he found that it was surrounded by four moons, which were describing orbits round Jupiter in very much the same way that our moon describes an orbit round the earth once in 27½ days. He was delighted because this observa-

tion is in some sense a confirmation of the Copernican theory of the solar system. It was soon found that these moons disappeared regularly behind Jupiter, showing that their orbits were in the same plane as the orbit of Jupiter round the sun. It was natural to try to find the time taken by each of these moons to go once round Jupiter and in making observations of this sort between 1672 and 1675 Römer noticed a rather interesting effect. Let us suppose that he was making observations when the earth and Jupiter were at their least distance apart, or in conjunction, as at E_1 and J_1 (Fig. 165) and that he was observing the rotation of one of the moons and found that it disappeared at 9.0 a.m. and again at 9.0 p.m. On repeating these observations, he found that the time at which he saw the moon disappear did not remain 9.0 a.m. or 9.0 p.m., but slowly yet systematically got later and later each time, until 0·545 of a year later, when the earth and Jupiter were at their greatest distance apart, or in opposition, as at E_2 and J_2, the moon disappeared some 16 minutes after 9.0 a.m. Continuing his observations, he found that in the next 0·545 year as the earth approached Jupiter again, the moon disappeared earlier than 9·16 a.m. and that its disappearance became earlier each day until when the earth and Jupiter were at E_3 and J_3 in conjunction once more the moon disappeared at 9.0 a.m. again. What is the interpretation of this regular variation in the time of the disappearance of the moon, or in the observed period of its rotation about Jupiter? It can be readily explained if we assume that light travels with a finite velocity. If this is so, then 9.0 a.m. is not the actual time at which the moon disappears for the conjunction E_1J_1, but it is the time at which the light signal from the moon recording the disappearance reaches the earth. Our news of the disappearance is late by the time light takes to travel the distance E_1J_1. It is now clear that as the earth recedes from Jupiter, the light signal has to travel further each day in order to deliver its message, and so the time at which it arrives gets later and later each day. It will be latest at an opposition when the earth and Jupiter are at the maximum distance apart, and the fact that it arrives about 16 min. later at an opposition than at a conjunction means that light takes this time to travel the distance E_2J_2–E_1J_1, which is the diameter of the earth's orbit round the sun.

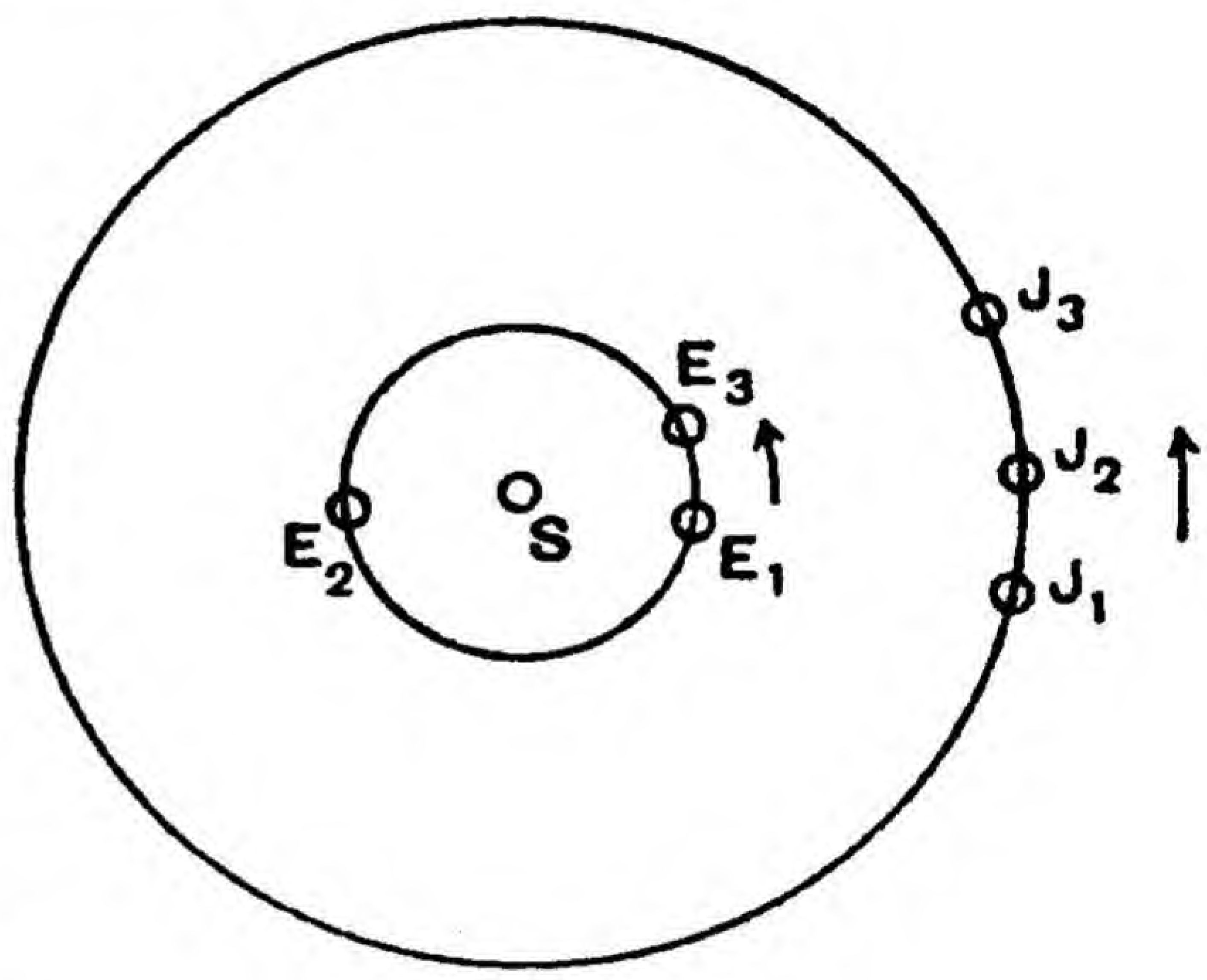


Fig. 165.

But we can now see that the above value of 12 hours for the period of the moon is not quite correct, as the earth will recede a small distance from Jupiter between the two disappearances, so that the above time of 16 min. is not quite the correct value of the time light takes to travel a diameter of the earth's orbit. We can calculate the true value in this way.

If the moon is observed to make n revolutions between a conjunction and an opposition and t is the true periodic time, D is the diameter of the earth's orbit round the sun and c the velocity of light, then the observed time of the n disappearances is $nt + \frac{D}{c}$. In the same way the observed time between the n revolutions occurring between an opposition and a conjunction is $nt - \frac{D}{c}$. These two times are measured and the difference between them is $\frac{2D}{c}$ and the most accurate value so far obtained is 32 min. 52 sec. So light takes 16 min. 26 sec. to travel once across a diameter of the earth's orbit, which is about 186,000,000 miles long. Thus the velocity of light is approximately 186,000 miles per sec. It is clear that the accuracy of this method depends not only on how accurately the above time can be determined, but also on the accuracy with which the diameter of the earth's orbit round the sun is known. Originally the above time and the diameter of the earth's orbit were used to find the velocity of light, but now the time and the velocity of light found by a terrestrial method are used to calculate the diameter of the earth's orbit. This important quantity is the astronomer's yard-stick, since it is used to measure the distances of the fixed stars and it is essential to know it as accurately as possible.

101. THE ABERRATION METHOD

If light travels with a finite velocity, it can be shown that the direction of a star as seen from the earth will be different according as the earth

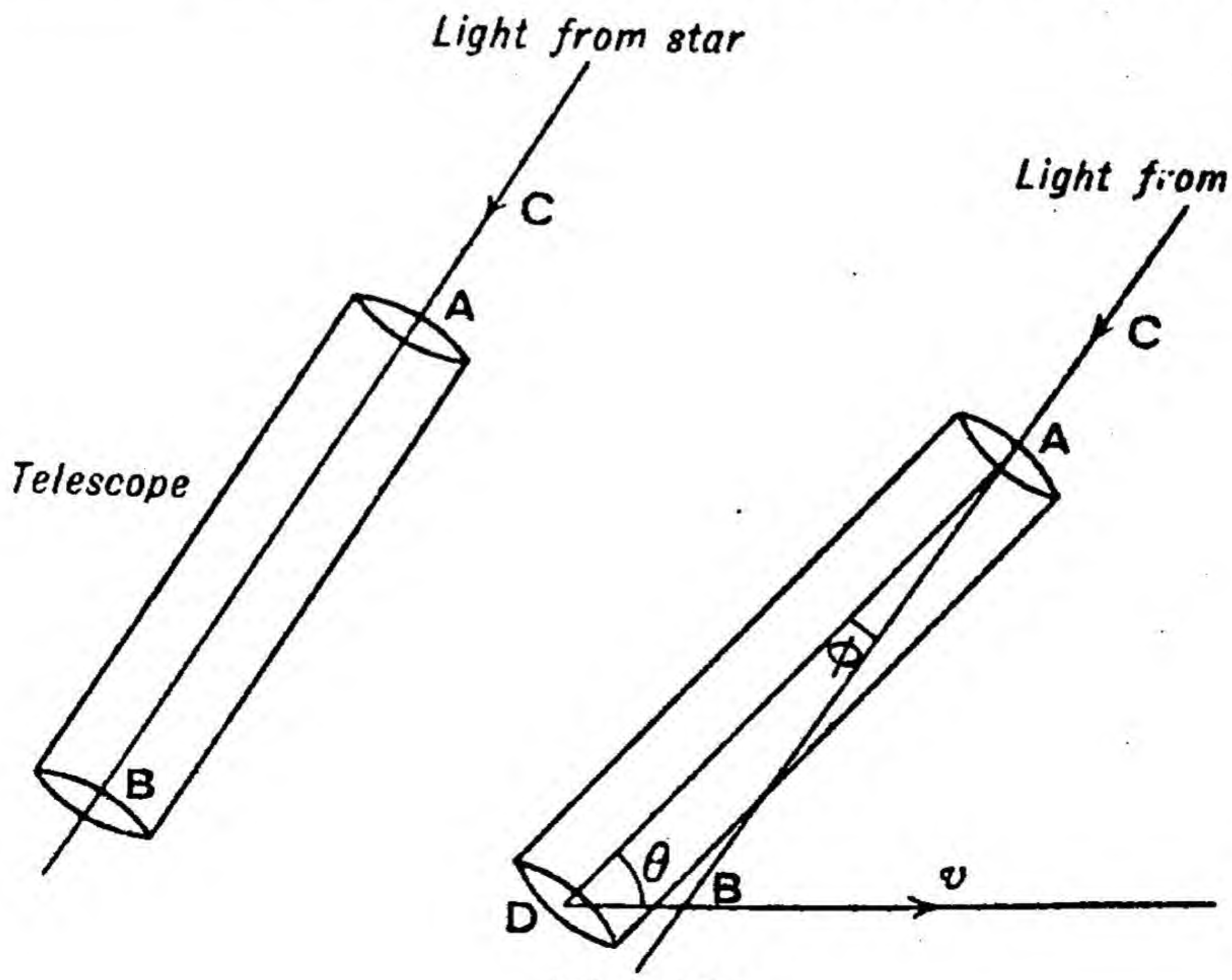


Fig. 166.

is moving directly towards the star or at a finite angle to the line joining the earth to the star. If the earth is moving directly towards the star, then it is evident that the optic axis of the telescope must be pointed in

the true direction of the star if the image of the star is to fall at the intersection of the cross-wires of the telescope (Fig. 166). But if the earth is not moving directly towards the star (Fig. 166), then the optic axis of the telescope must be set along DA in order to see a star whose true direction is BA, where DB is the distance which the earth, and therefore the telescope, travels in the time in which light travels AB. Thus if the light enters the telescope objective at A moving in the direction AB, by the time it reaches B the centre of the eyepiece will have moved from D to B, so that the ray will appear to an observer on the earth to have entered and emerged from the centre of the telescope and the image of the star will fall on the intersection of the cross-wires. The star will appear to be in the direction DA, an angle ϕ in front of its true direction. This angle is called the **aberration** of the star, and from the triangle ABD we have

$$\frac{DB}{\sin\phi}=\frac{AB}{\sin\theta}$$

$$\therefore \sin\phi=\frac{DB}{AB}\sin\theta$$

$$\therefore \phi=\frac{v}{c}\sin\theta$$

where v is the velocity of the earth and c the velocity of light and $\sin\phi$ is written as ϕ, since ϕ is always a few seconds of arc. The aberration is a maximum when the angle between the direction of motion of the earth and the apparent direction of the star is 90° and zero when this angle is zero. Therefore if a star is in the same plane as the earth's orbit round the sun, it appears to an observer on the earth to describe a small straight line in the heavens once in each year (Fig. 167), being in its true direction at the two positions E_1 and E_3 when the earth is approaching or receding directly from it and being furthest from its true direction at the two intermediate positions E_2 and E_4 when the earth is moving at right angles to its apparent direction. If the star is in a direction normal to the plane of the earth's orbit, it appears to describe a circle whose diameter subtends the same angle at the earth as the length of the line in the previous case. The value of this angle is $\frac{2v}{c}$. Stars in an intermediate position describe an ellipse.

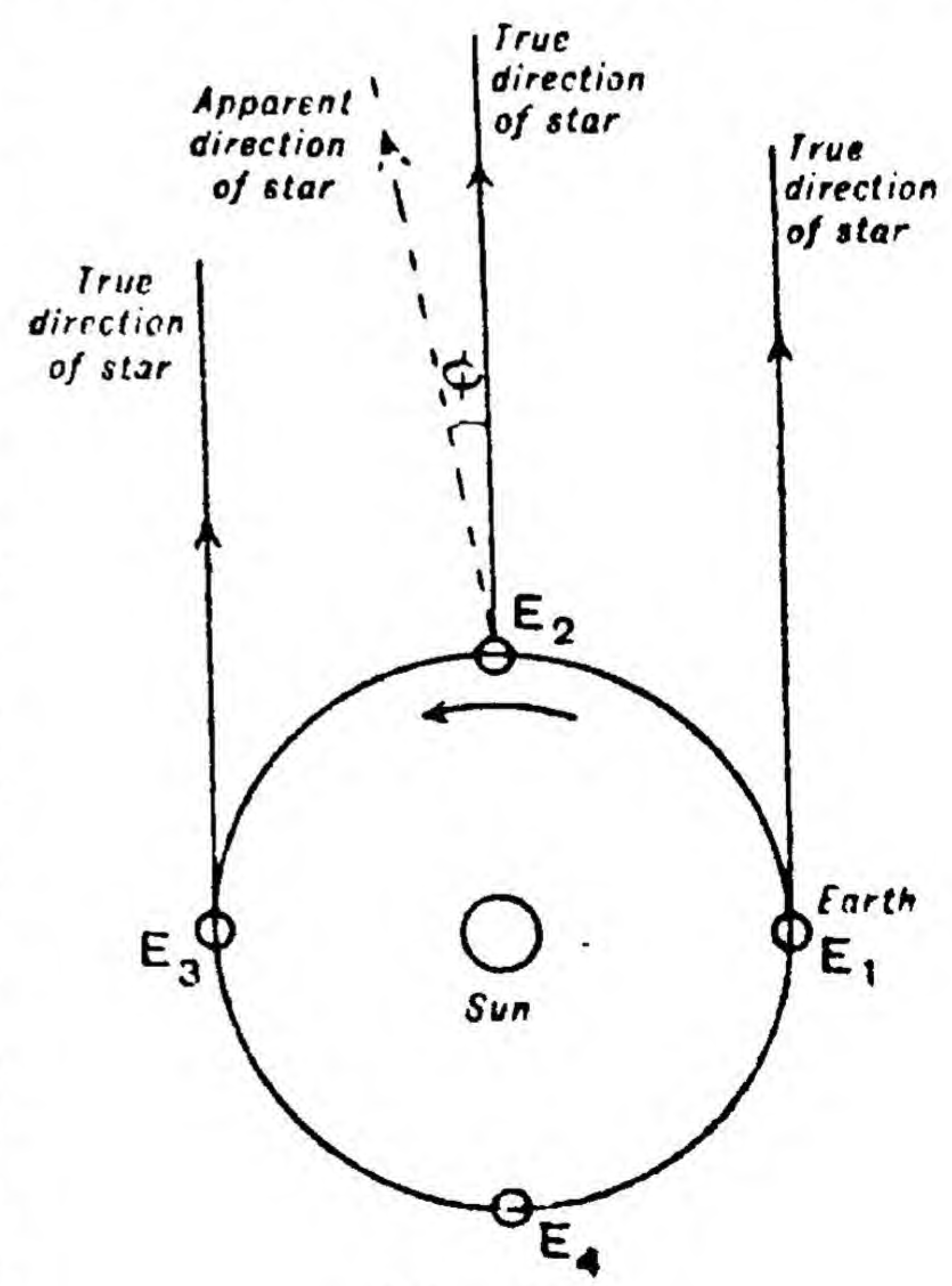


Fig. 167.

This effect was first observed by Bradley in 1727. At that time it was believed that the sun was the centre of the solar system and that the earth was describing an orbit round the sun. If this is the case, the near stars should move to and fro against the background of the more distant ones, in just the same way that near objects move to and fro against the background of the distant hills to a child on a swing. This motion is called parallax and the motion of the near object is in the opposite direction to that of the observer. It was the absence of this parallax which led the Greeks to reject the idea of a moving earth, but the work of Copernicus and others seemed to suggest the motion of the earth so strongly that it occurred to astronomers that the absence of parallax was due to the fact that the stars are so far away compared with the distance from the earth to the sun that the amount of the parallax is very small. When Bradley started to look for parallax, using a telescope, he was surprised and puzzled when he found, not a motion of the star in the opposite direction to that of the earth but in the same direction! On further investigation, he found that the maximum angle of displacement of the star from its mean position was the same for all stars, and this suggested that the effect was due to the motion of the earth alone. He then hit on the above explanation and saw that his observation could be used to measure the velocity of light. The best value of the maximum value of the angle of aberration so far obtained is 20·4 sec., and from this we have

$$\frac{v}{c}=\frac{20{\cdot}4\times\pi}{3600\times 180}$$

Taking the velocity of the earth in its orbit round the sun as 18·5 miles per sec., we have $c=187{,}000$ miles per sec.. It will be seen that this method also involves a knowledge of the diameter of the earth's orbit round the sun, and it is now used to calculate this diameter rather than to find the velocity of light.

102. FIZEAU'S TOOTHED-WHEEL METHOD

It is clear that we need to devise terrestrial methods if we are ever to find the velocity of light in water and to be independent of a knowledge

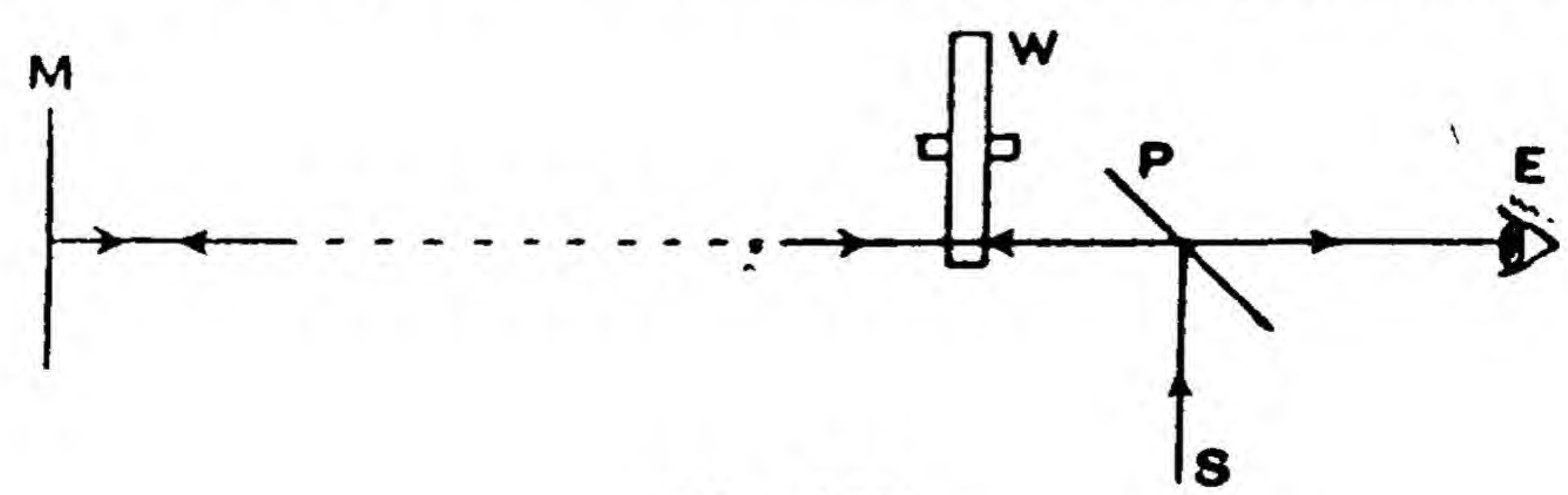


Fig. 168.

of the earth's distance from the sun. Is it possible to improve on Galileo's attempt in the light of modern technique so as to get a positive result from it? It is clear that the first thing is to replace the observer B (Art. 99)

by a mirror, and we must have at the sending end some mechanism capable of measuring a time interval of 1/20,000 sec., which is the time taken by light to travel some 10 miles from the above values of the velocity of light. In 1849 Fizeau described a way of doing this. The principle of his method is shown in Fig. 168. A source of light S sends a ray of light on to a glass plate P, from which some is reflected to a mirror M about 5 miles away, passing through one of the spaces of a toothed wheel W on the way. The ray of light is reflected by M back through the toothed wheel W and the glass plate P to the eye at E. If the wheel is rotated so fast that one space is succeeded by the next tooth in the time the light takes to travel from W to M and back, it is clear that the observer at E will not see an image of the source S. If the number of teeth of the wheel and its rate of revolution are known, it is clear that the velocity of light can be calculated.

In practice it is useless to work with a single ray of light, as the final image would be too faint, and a beam of light must be used, the apparatus being arranged as shown in Fig. 169. The rays from the

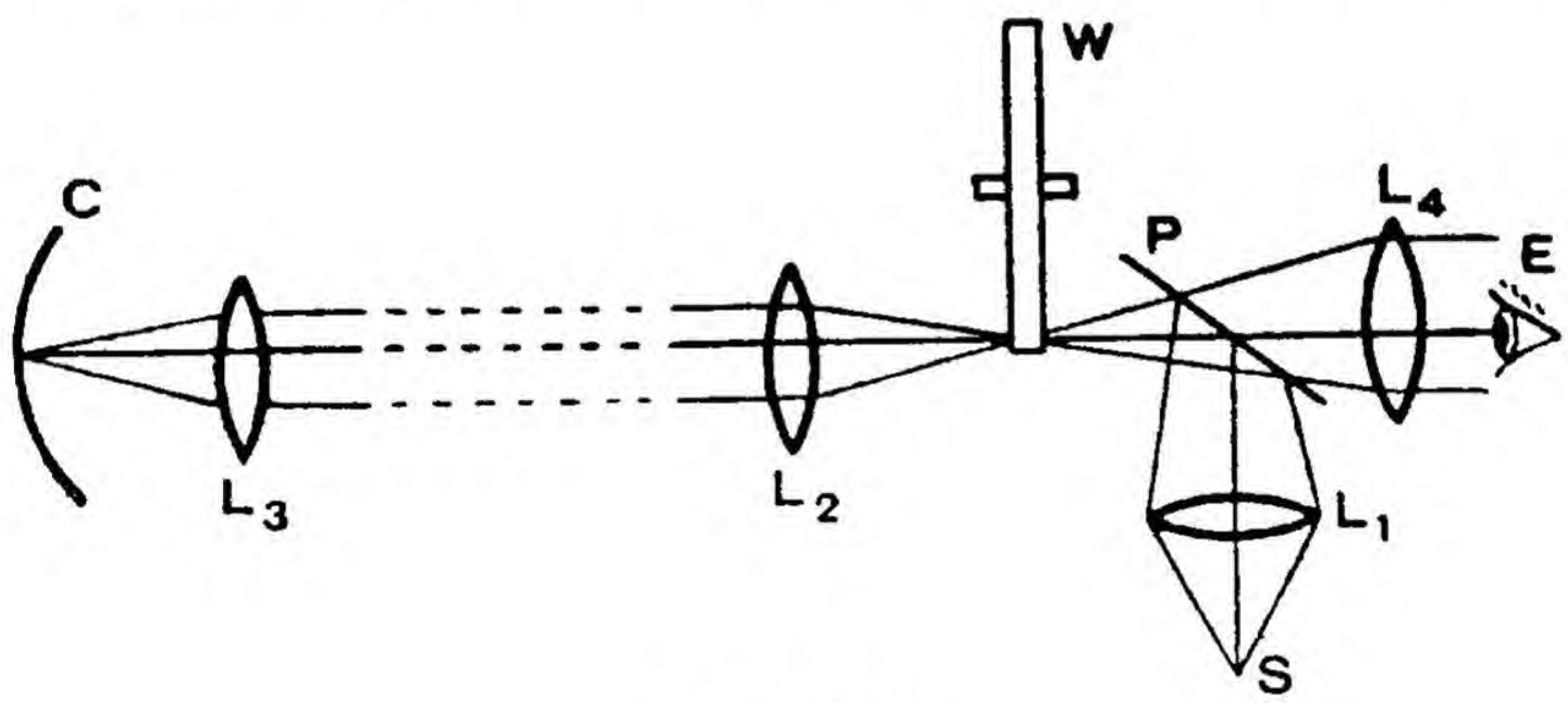


Fig. 169.

source of light S, an arc lamp, are focussed by the lens L_1 and reflected by the glass plate P so as to come to a point at one of the spaces of the toothed wheel W. The light diverges to the lens L_2, which makes it parallel, after which it proceeds to the lens L_3 some 5 miles away. It is brought to a focus by this lens at the surface of a concave mirror C, whose centre of curvature coincides with the centre of the lens L_3. This adjustment is most important, since it ensures that the central ray of the beam retraces its incident path after reflection and so it and the whole beam strike the lens L_2 on the return journey. The reason for this is as follows. The central ray passes through the centre of the lens L_3, which is also the centre of the curvature of the mirror C, so it strikes the mirror normally and retraces its path to the centre of the lens L_3, through which it passes undeviated. After that it retraces its incident path to the lens L_2. Since the beam as a whole is brought to a focus at the surface of the mirror, this changes the point image to a point object in the focal plane of the lens L_3. Since this point object coincides with the point image of the incident beam, the reflected beam on emerging

from L_3 must coincide in direction *and position* with the beam incident on L_3 from L_2 and so this reflected beam will strike the lens L_2 once more. It is brought to a focus at one of the spaces of the toothed wheel again and passes through the glass plate P and on to an eyepiece L_4, through which an image of the source is viewed by the eye at E.

If the wheel is rotated slowly, an image of the source is seen intermittently, but this becomes a steady image when more than sixteen spaces pass the beam of light every second owing to the persistence of vision. But if the speed of rotation of the wheel is increased until one space is succeeded by the next tooth in the time the light takes to travel from W to C and back again, the image of the source will disappear. If the speed of the wheel is doubled, the image will be at its greatest brightness again, since the light which gets through one space on its way out to C will get through the next space on its way back to the eye. If the speed of rotation of the wheel is increased to three times its value at the first disappearance, the second disappearance will occur when the light passing through one space on the outward journey is stopped by the next tooth but one on the return journey. In this way, many appearances and disappearances can be observed.

In an actual experiment the apparatus is set up and the speed of rotation of the wheel is adjusted until the r^{th} disappearance occurs. It is evidently possible to adjust the wheel more accurately for a disappearance than for a maximum brightness, since it is easier to distinguish between illuminations of 0 and 1 than of 99 and 100. The number of revolutions n of the wheel per second is then measured, and the distance d from the toothed wheel to the concave mirror and the number of teeth N of the wheel are known. Then light travels a distance $2d$ in the time one space is succeeded by the next tooth but $r-1$, or in the time that $r-\frac{1}{2}$ cogs pass the beam of light. Thus light travels a distance $2d$ in $\frac{r-\frac{1}{2}}{nN}$ secs. So the velocity of light c is given by

$$c=\frac{2d}{\dfrac{r-\frac{1}{2}}{nN}}$$

$$\therefore\ c=\frac{4nNd}{2r-1}$$

In one of Fizeau's experiments d was 8·6 Km., N was 720. It is interesting to calculate the value of n for the first disappearance, assuming c to be 300,000 Km. per sec. It works out at 12·1 revs. per sec., quite a possible value.

The reader may ask how the apparatus is actually set up. How is it possible to make the parallel beam from L_2 fall on the lens L_3 5 miles away ? The wheel W is removed and replaced by cross-wires and the

concave mirror is similarly replaced by cross-wires and an eyepiece is placed on the side of L_3 remote from L_2 so that L_3 and this eyepiece act as an astronomical telescope. This telescope is pointed at the apparatus 5 miles away and it is moved about until the lens L_2 and the cross-wires are picked up. It is then adjusted until the cross-wires replacing W coincide with those replacing C, when the optical axis of L_3 is in the same straight line as that of L_2. The cross-wires are removed and the wheel W is placed in position at the observing end of the apparatus, while the concave mirror is carefully adjusted so that its centre of curvature coincides with the centre of the lens L_3, when the apparatus is ready for use.

The apparatus has been improved by later workers, notably Cornu and Young and Forbes. It is evident that, at a disappearance, some of the light strikes the teeth on its outward journey and is reflected straight back into the eyepiece. As this happens many times a second, it produces a faint uniform illumination, which makes it difficult to decide exactly when the image has disappeared. This is avoided by bevelling the teeth, to reduce this reflection to a minimum. The intensity of the image is increased by silvering the plate P, leaving a small hole in the middle through which light can pass to the eye at E. It is also possible to change the speed of rotation of the wheel by a finite amount while the image remains absent and so it is difficult to find the speed at which the returning light is stopped by the middle of a tooth. This difficulty was avoided by measuring the speed of rotation of the wheel for equal brightness of the image just before and just after a disappearance and taking the mean of these speeds as that required to produce an exact disappearance. Finally the distance d was increased to 29 miles.

103. FOUCAULT'S ROTATING MIRROR METHOD

Fizeau's method cannot be used to measure the velocity of light in water, since a column of water some 5 miles long would absorb the light completely, but Foucault finally measured the velocity of light

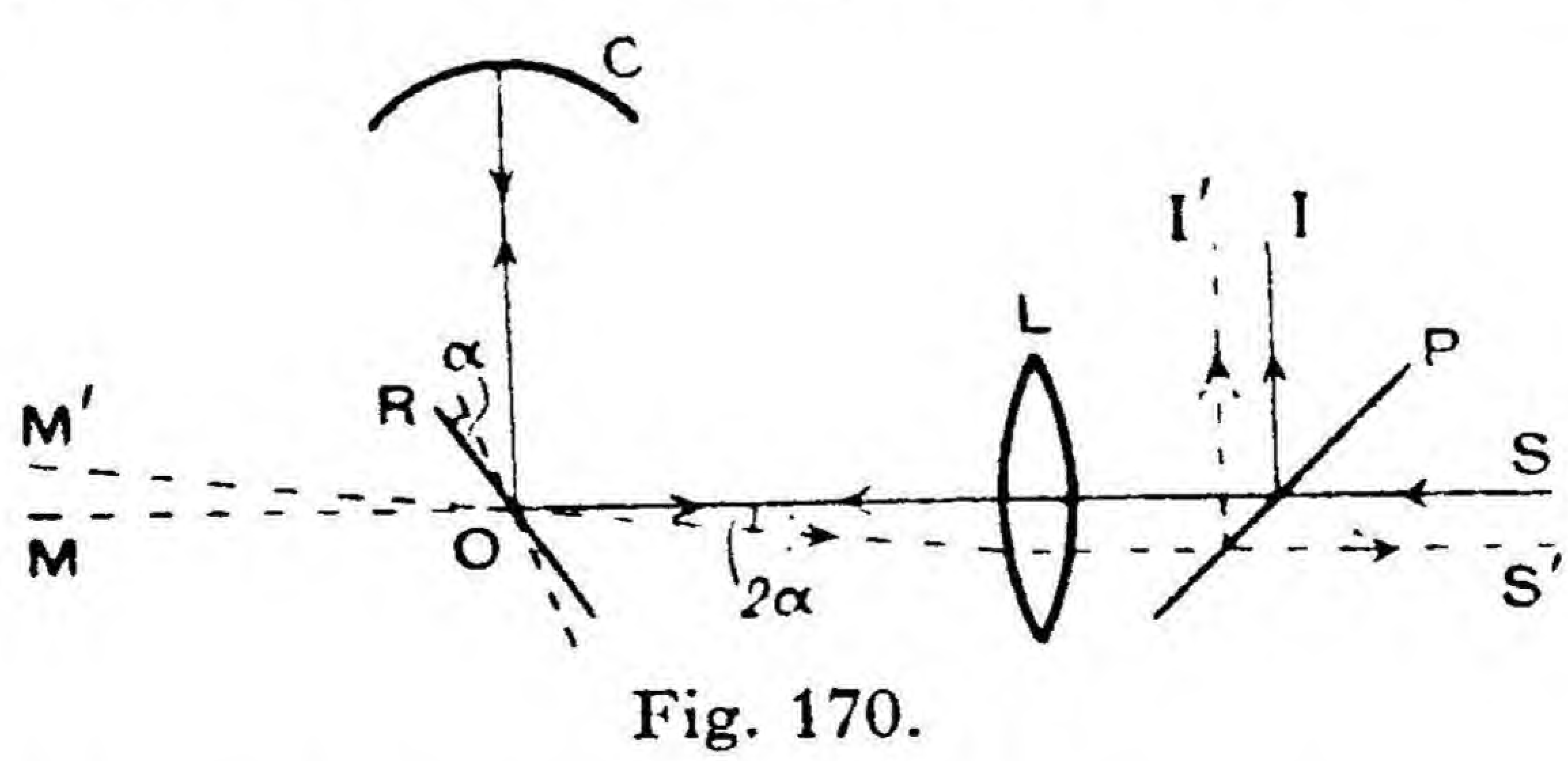


Fig. 170.

by a method which could be used to find that in water. The principle of his method is illustrated in Fig. 170, in which only the central ray of the beam is shown for simplicity. Light from a source S passes

through a glass plate P and a lens L on to a plane mirror R, which reflects it on to a concave mirror C. Since the centre of curvature O of the mirror C coincides with the axis of rotation of the rotating mirror R, the ray strikes the mirror normally and so retraces its path right back to the plate P, where some of it is reflected to form an image of the source S at I. If the mirror is rotated fast enough, it will turn through an angle α in the time the light takes to travel from R to C and back again, so that the ray reflected from R will make an angle 2α with its original direction and the final image will be deflected to I′. If the deflection of the image due to the rotation of the mirror is measured, this angle can be calculated and, knowing the speed of rotation of the mirror, the time taken for the light to travel from R to C and back can be found and then the velocity of light.

The apparatus is shown more fully in Fig. 171, where it is seen that the lens L forms a real image of the source S at the surface of the concave mirror C. The mirror R reflects the light from this image so that it

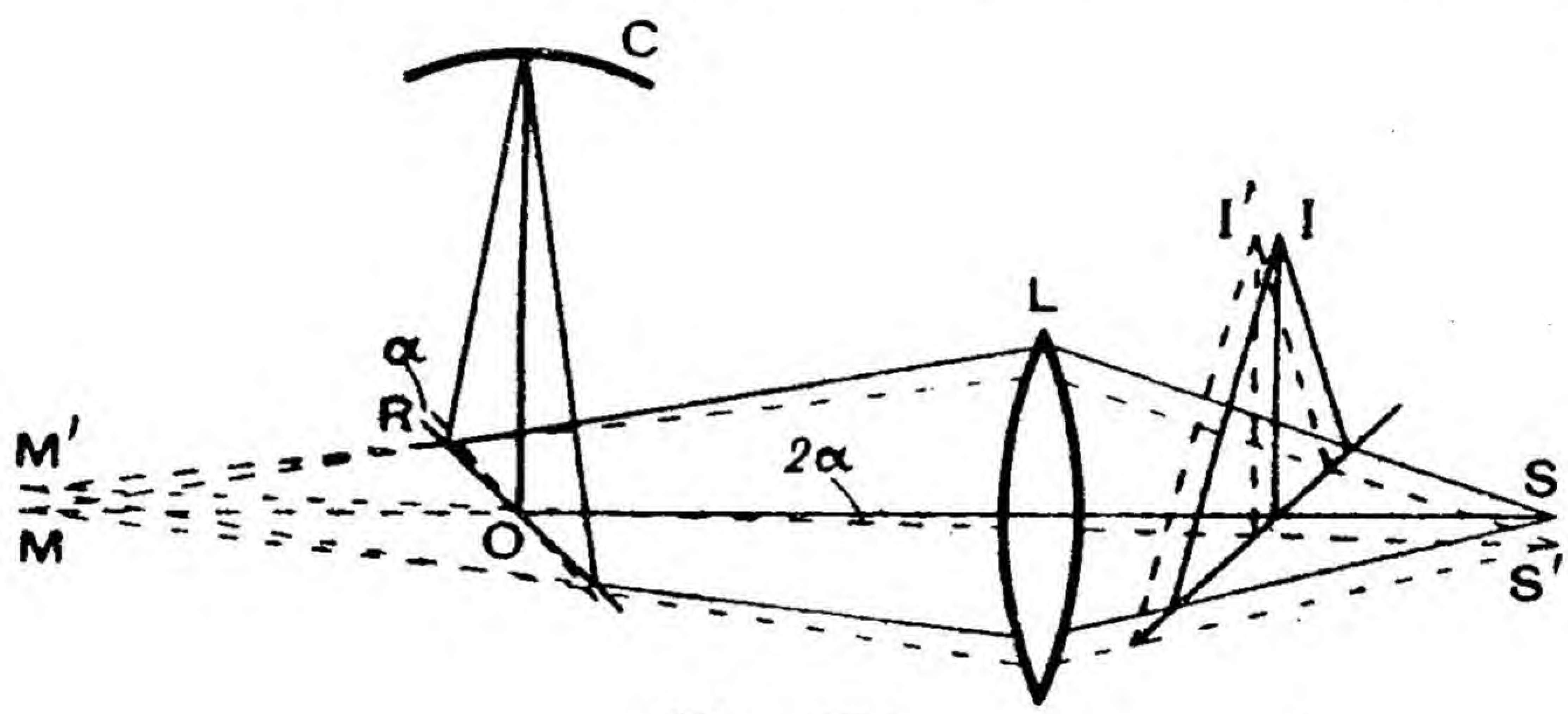


Fig. 171.

appears to diverge from M and the lens L and plate P form a real image of M at I, which is observed with a micrometer eyepiece. Since the centre of curvature of the concave mirror and the axis of rotation of the rotating mirror coincides, the central ray of the pencil always strikes the concave mirror normally and retraces its path, so the image M from which rays appear to come after the second reflection from the rotating mirror is always in the same place. Therefore *the position of the final image* I *is independent of the position of the rotating mirror* always provided that it sends some light on to the concave mirror. When the mirror is rotated slowly, the image is not drawn out into a band, but remains fixed in position, but when the mirror is speeded up so that it rotates through a finite angle α while light travels from R to C and back, the image M is displaced to M′ and I to I′, the angle MOM′ being 2α.

In an actual experiment the micrometer eyepiece is adjusted so that the image I falls on the intersection of the cross-wires and the mirror R is rotated at a constant rate. The eyepiece is then adjusted so that the displaced image I′ falls on the centre of the cross-wires and the deflection t of the image is measured. The distance d from the rotating to the

concave mirror is known and the number of revolutions n of the mirror in one second are measured. Light travels a distance $2d$ in the time the mirror takes to rotate through an angle α, which is $\frac{\alpha}{2\pi n}$. So the velocity of light c is given by

$$c=\frac{4\pi nd}{\alpha} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

But α must be calculated from the deflection of the image and the distances a and b of O and S respectively from the centre of the lens L. Now

$$\alpha=\tfrac{1}{2}\angle\mathrm{MOM}'=\tfrac{1}{2}\frac{\mathrm{MM}'}{d}$$

Also $$\frac{\mathrm{MM}'}{a+d}=\frac{\mathrm{SS}'}{b}$$

$$\therefore\ \alpha=\tfrac{1}{2}\frac{(a+d)\mathrm{SS}'}{bd}$$

But $$\mathrm{SS}'=\mathrm{II}'=t$$

$$\therefore\ \alpha=\tfrac{1}{2}\frac{(a+d)}{db}t$$

Substituting this value of α in equation (56) we have

$$c=\frac{8\pi nbd^2}{(a+d)t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

In Foucault's experiment d was 20 metres, and the mirror was rotated fast enough to get a deflection of the image of 0·7 mm. Taking b as 1 metre, $a+d$ as 21 metres, and c as 3×10^8 metres per sec., we get n to be 438 revs. per sec., which it is quite possible to produce and maintain constant.

Foucault's method is admirably adapted to seeing if the velocity of light is greater in air than in water, for it is only necessary to insert a tube of water between the rotating and concave mirrors and see if the deflection increases or gets less. Foucault did this and found that the deflection increases and later Michelson made actual measurements of the deflection and showed that the ratio of the velocity of light in air to that in water is 1·33. Therefore light travels more slowly in water than in air, which agrees with the wave theory (Art. 96). Does this mean that the corpuscular theory is disproved? We shall discuss this question at the beginning of the next chapter.

104. IMPROVEMENTS OF FOUCAULT'S METHOD

Returning now to the accurate determination of the velocity of light in air, we see that Foucault's method is not capable of great accuracy since the deflection of the image is only 0·7 mm. Foucault made this

as accurately measurable as possible by using as his source a narrow slit, but it is clearly necessary to increase the deflection to make the method more accurate. This can be done both by increasing the distance from the source S to the lens L and also the distance d. But an increase in d decreases the brightness of the image I′, since light only goes to form this image for that fraction of a revolution of the rotating mirror, during which the beam from it is striking the concave mirror and this fraction decreases as the distance d increases. It should be emphasised that this brightness of the image I′ is a serious matter, since it is seen against a background of steady illumination produced by the light reflected directly from the rotating mirror into the lens L and to the eyepiece. How can the brightness of the image be kept constant while increasing the distance d? Michelson saw that it could be done by placing the lens L between the rotating mirror and concave mirror and making it of such a long focal length that the image of S still came at the surface of the concave mirror (Fig. 172). S_1 and S_2 are the images of S formed by the rotating mirror in the positions R_1 and R_2 respectively, the extreme positions in which the rotating mirror sends light on to the concave mirror. As the rotating mirror describes one revolution, the image of S in it describes a circle centre O radius a_1 and while it describes the arc of length S_1S_2, it is contributing light to form the image I′. If S_1S_2 is kept constant for a given diameter C_1C_2 of the concave mirror, the brightness of the image will remain the same. Now

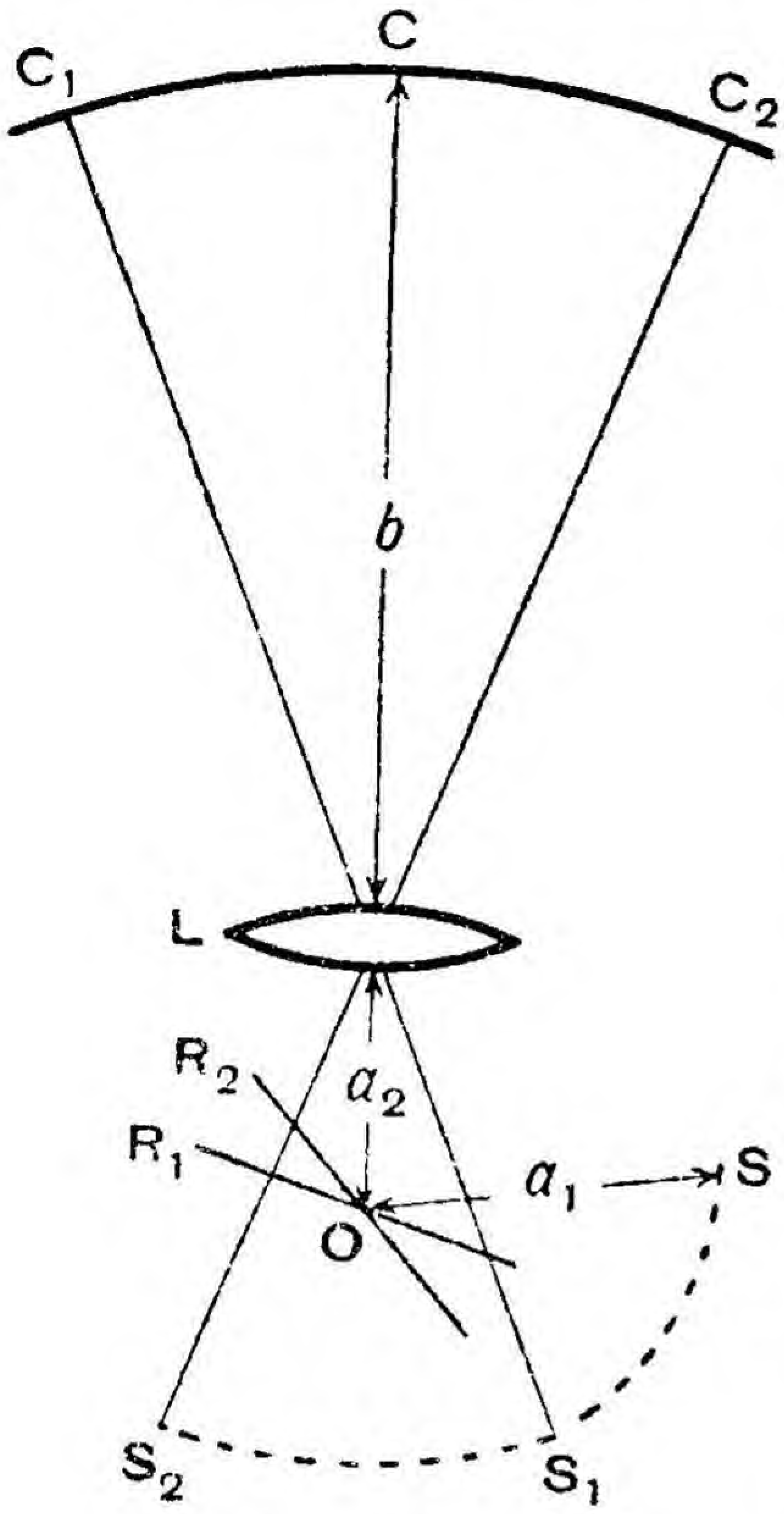


Fig. 172.

$$\frac{S_1S_2}{C_1C_2} = \frac{a_1+a_2}{b}$$

So S_1S_2 can be kept constant for a given value of a_1 by increasing a_1+a_2 and b in such proportions as to keep $\frac{a_1+a_2}{b}$ the same. This can be done by increasing the focal length of the lens L so that the above ratio remains the same while the total distance from the image to the object increases. In this way Michelson was able to increase the distance d to 700 metres and obtain a deflection of the image of 133 mm.. The brightness of the image can be still further increased by using a four-sided mirror instead of a mirror silvered on one face only, and this improvement was first introduced by Newcomb. To avoid an equal increase in the brightness of the background a long rotating mirror was used, and arrangements were made so that the light was incident on its lower half and was finally

reflected off the upper half back into the micrometer eyepiece. Therefore light reflected directly from the rotating mirror never entered the eyepiece.

105. MICHELSON'S NULL METHOD

The most accurate method of measuring the velocity of light which has so far been developed is an extension of the four-sided mirror method of Newcomb. It is well known that it is possible in all physical measurements to decide more accurately that a given quantity is zero than to measure it. For this reason many measurements of the electrical quantities such as P.D., resistance, and capacity are done by making adjustments until the deflection of a galvanometer is zero. Newcomb suggested that the mirror should be rotated at such a speed that one face exactly succeeds the next while the light travels from the rotating mirror to the concave mirror and back, when the deflection of the image would be zero. Assuming that the mirror can be rotated at 1000 rev. per sec., light can still travel some 46 miles while one face of a four-sided mirror is taking the place of the next, so Michelson's 700 metres must be increased to some 23 miles ! Michelson started in 1924 to try to realise Newcomb's suggestion and he succeeded by using eight, twelve, and sixteen sided mirrors over a base-line 22 miles long between Mount Wilson and Mount San Antonio in California.

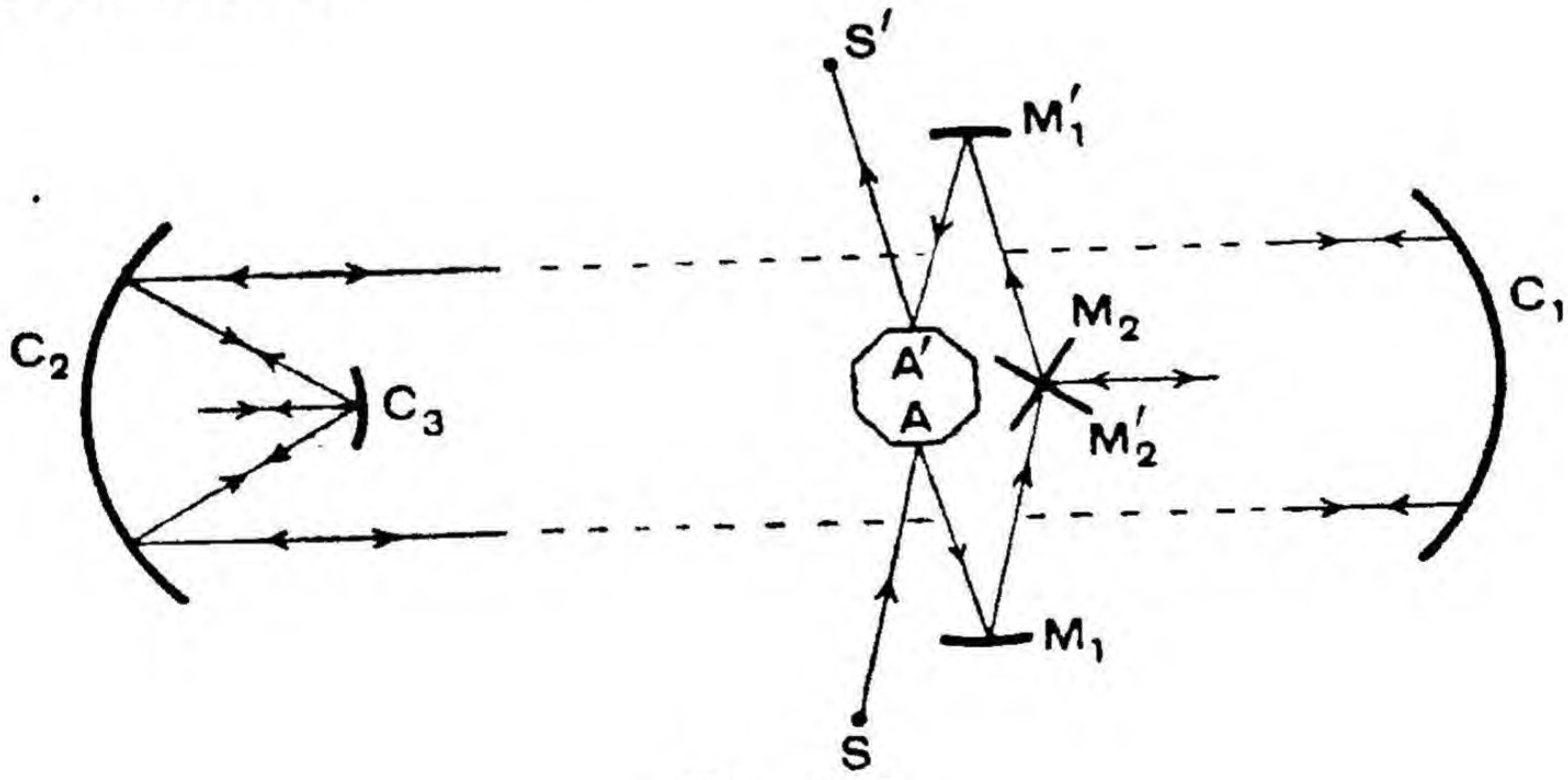


Fig. 173.

The apparatus which he used is represented diagrammatically in Fig. 173. Light from the source S strikes the lower part of the face A of the rotating mirror and is reflected from mirrors M_1 and M_2 on to the concave mirror C_1, from which it travels to the distant station at Mount San Antonio striking the mirror C_2 and being brought to a focus at the surface of another concave mirror C_3. Exactly as in Fizeau's experiment, it retraces its path and is reflected from the mirror M'_2, which is at right angles to and above M_2. It then passes to M'_1, which reflects it to the upper part of the face A' of the rotating mirror to form a final image at S', which is observed through a micrometer eyepiece. The return ray is reflected

from a different face of the rotating mirror from the outgoing ray so as to prevent scattered light entering the eyepiece.

When the apparatus has been set up, the eyepiece is adjusted so that the image S′ falls on the intersection of the cross-wires and the mirror is set rotating and its speed is adjusted so that the image is undeflected. This means that one face of the rotating mirror exactly succeeds the next while the light travels from A to the distant mirror and back to A′. This distance was measured to 1 in 7,000,000 and the speed of rotation of the mirror was measured by a stroboscopic method involving ultimate comparison with a standard clock. The angle between the faces of the mirror was 45° in the case of the eight-sided mirror to 1 in 1,000,000. If the mirror makes n revolutions per second for zero deflection of the image and d is the total distance travelled by the light on the outward journey, $c = 16\,nd$ for an eight-sided mirror.

Michelson, Pease, and Pearson started in 1930 to improve on these experiments by working with a base a mile long enclosed in a tube which was exhausted to 0·5 mm. of mercury so that they could measure the velocity of light in a vacuum directly. After Michelson's death in 1931 Pease and Pearson continued the work alone. They used a 32-sided mirror and revolved it at such a speed that there was no deflection of the final image formed by light reflected nine times up and down the base line. These experiments by the null method are the most accurate determination of the velocity of light which has so far been made and the results are shown in Table 13. A little reflection will show that the principle of the null method is the same as that of Fizeau's method, the many-sided revolving mirror replacing the toothed wheel and the zero deflection of the image being a very accurate way of deciding when one face has succeeded the next, which is much better than trying to decide when one space of the toothed wheel has exactly succeeded the next by seeing if the brightness of the image is a maximum. Thus the latest and most accurate determination is just Galileo's method refined by modern technique !

TABLE 13

Type of Mirror.	Velocity of Light in Vacuum in Kms. per sec.
8-sided glass.	299,797
8-sided steel.	299,795
12-sided glass.	299,796
12-sided steel.	299,796
16-sided glass.	299,796

106. THE VARIATION OF THE VELOCITY OF LIGHT WITH TIME

We shall conclude this account of measurements of the velocity of light by reference to a possible variation with time which is suggested

by an examination of the results obtained in the last sixty years. They are shown in Table 14, in which the results are given as the velocity in vacuum. This is because the refractive index of air relative to a vacuum is 1·00028, and so light travels 1·00028 times more quickly in a vacuum than in air. It is desirable to express the results as the velocity in a vacuum, since its properties are invariable. The results shown in Table 14 suggest a small variation of period forty years, but much more work is required before this tentative conclusion can be accepted. In order to do this, it is necessary to devise a method which can be left set up in a laboratory so that measurements can be made at regular intervals of, say, a year. The method must be as accurate as the null method if the variation is to be detected with certainty. Is it possible to do this ? Can the toothed wheel be rotated so fast in Fizeau's method that the base can be shortened to a few metres instead of several kilometres ? It seems hopeless to do this with the inertia of a material wheel, but a lot has been done in recent years in oscillographs by replacing the material moving parts by the electron with its negligible inertia. So we should look for the solution of this problem along such lines as these, paying particular attention to any phenomenon likely to be suitable for the purpose.

TABLE 14

Date.	Investigator.	Velocity in Vacuum in Kms. per sec.
1874	Cornu.	299,990 ± 200
1879	Michelson.	299,910 ± 50
1883	Newcomb.	299,860 ± 30
1883	Michelson.	299,853 ± 60
1902	Perrotin.	299,901 ± 84
1926	Michelson	299,795 ± 4
1932	Pease and Pearson	299,774

EXAMPLES ON CHAPTER XI

1. Describe Römer's method of measuring the velocity of light. State and explain at what positions of the earth in its orbit round the sun the time between two successive eclipses of one of Jupiter's moons is (a) greatest, (b) least. Calculate the value of each of these times in the case of the moon whose period round Jupiter is exactly seven days to an observer on Jupiter. The velocity of light is 186,000 miles per second and the velocity of the earth in its orbit round the sun is 18·5 miles per second.

2. Prove that the finite value of the velocity of light causes a star in the plane of the earth's orbit round the sun to appear to describe a periodic motion of period one year along a line, while a star whose direction from the earth is normal to the plane of the earth's orbit round the sun appears to describe a circle in a time of one year. Show how this apparent motion of the stars can be used to measure the velocity of light.

3. Discuss Fizeau's toothed-wheel method of measuring the velocity of light. If we wished to make a determination by this method at Repton, it would be possible to get two stations about 7 miles apart. If we used a cog-wheel with thirty teeth from the gear-box of an old motor car as the toothed wheel, at what rate should we have to rotate the wheel for the first disappearance? Do you regard the attempt to be worth making?

4. Describe the revolving mirror method of determining the velocity of light, and show how the result is calculated from the observations made. (*O. and C.*)

5. Describe how the velocity of light has been accurately determined by a method based on reflection from a rotating mirror. If the experiment you describe is carried out with water in the path of the light, what results are obtained and what is their theoretical significance? (*London.*)

6. In Foucault's rotating mirror method, a concave mirror of radius of curvature 10 metres is used, the illuminated slit is 1·5 metres from the lens, and the distance from the lens to the surface of the concave mirror is 11·5 metres, these distances being measured along the central ray of the beam. If the rotating mirror makes 800 revolutions per second, find the deflection of the image, given that the velocity of light is 3×10^{10} cm. per second. Work from first principles or prove the formulæ you use.

Suppose the lens is removed and only the central ray is used, find the displacement of the image now.

7. How has the velocity of light been measured? How did measurements of the velocity of light support the wave theory of light propagation? (*London Inter.*)

8. How has the velocity of light been determined by a terrestrial experiment? Explain the refraction of light from air to water in terms of (a) the emission (or corpuscular theory) of light, (b) the wave theory of light. What evidence was given on this matter by measurements of the velocity of light in water? (*N.U.J.B.*)

9. Discuss the difficulties in Foucault's rotating mirror method and describe fully how they have been overcome in Michelson's modification of it.

10. Describe a method of measuring the velocity of light. Indicate how the speed of sunlight varies in the course of its passage from the sun to the bottom of a lake on the earth's surface. (*London.*)

11. Describe a method for the accurate measurement of the velocity of light. How has the velocity of light been measured (a) in air, (b) in water? What is the theoretical importance of the results obtained? (*Camb. Schol.*)

12. Describe the methods by which the velocity of light has been determined. The velocity of light in water (refractive index 1·33) is $2{\cdot}55 \times 10^{10}$ cm./sec. In carbon bisulphide (refractive index 1·64) it is $1{\cdot}705 \times 10^{10}$ cm./sec. Comment on these figures. (*Oxford Schol.*)

13. Give a critical discussion of Michelson's null method of measuring the velocity of light. If a base line of 7 miles and a 24-sided mirror are available, do you consider that a determination by this method is feasible?

14. Discuss the results which have been obtained from determinations of the velocity of light. What further work does your discussion suggest?

Chapter XII

THE PROPERTIES AND THEORY OF WAVES

107. INTRODUCTORY

We have just seen that, when the velocity of light was measured in water, it was found to be less than that in air. This result is in accordance with the prediction of the wave theory but contrary to that of the corpuscular theory. Does it mean that this theory must now be abandoned? Not at all! The supporters of the corpuscular theory will rightly contend that it would be equally reasonable to abandon the wave theory on account of its inability to explain rectilinear propagation. And they may add that it is quite possible to explain the above experimental fact on the corpuscular theory. It is only necessary to assume that both the tangential and normal components of the velocity of the corpuscle decrease as it goes from air to water through a narrow region on either side of the refracting surface, and that the tangential component suffers that greater decrease than the normal component which is needed to make the ray bend by the observed amount towards the normal. The supporters of the wave theory may object that this is precisely the kind of *ad hoc* assumption which is to be avoided, since it cannot be made until after the fact is known and because a theory which has to resort to *ad hoc* assumptions cannot suggest new lines of investigation. But those who believe in the corpuscular theory will reply that the adherents of the wave theory should be the last persons to complain about *ad hoc* assumptions, since they have filled the whole of space with a medium, which has never been detected, merely to explain that light can travel through empty space! We cannot, then, regard the fact that the velocity of light in water is less than that in air as overthrowing the corpuscular theory; both theories are still in difficulty and the acid test is still to come. It is this: which of the two will prove the more fruitful in suggesting new lines of investigation? Can the corpuscular theory suggest any properties which light should have which have not yet been discovered? What can the wave theory do in this respect? There are three ways in which it can suggest further advance.

We have seen that the known waves obey roughly the same laws of reflection and refraction as light. But it is time to go beyond mere similarity and to proceed to construct a mental picture of waves, which will be expressible in mathematical terms. We must make a kind of

working thought model of waves ; our model will probably be rather simpler than the actual water and sound waves, but it has the advantage that we can deduce logically what waves will do and extend our ideas to a much wider range of conditions than we have so far observed in practice. We must extend this model to explain reflection and refraction. There are two other well-known phenomena exhibited by water and sound waves, namely, interference and diffraction. We must therefore go into these phenomena and see that they follow from our theory and then we must try to obtain the corresponding phenomena in the case of light. The wave theory, then, does seem to offer some promising lines of investigation ; let us occupy ourselves with a general theory or model of waves and investigate interference and diffraction in water waves and sound ; then we shall be ready for a further step into the unknown.

108. WAVES

A wave is the propagation of a condition through a medium without any translation of the medium in the direction of the waves. The word " translation " means that there is no bodily motion of the medium, no transference of the medium in the direction of the waves. For example, if a rope is held in the hand (Fig. 174) and the free end is moved up and down once, a hump is produced in the rope

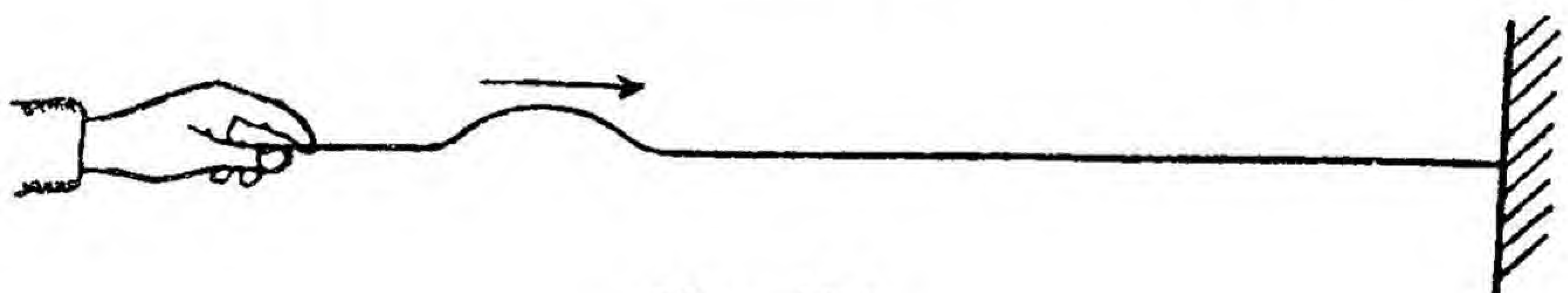


Fig. 174.

and this hump, or crest, travels along the rope at a definite velocity. This is a wave and it is to be emphasised that it is the crest, a condition of the rope, which travels along and not the rope itself. Again, if a plank floating on the surface of a pond is moved up and down, crests and troughs are produced and travel over the surface of the water. In this case the motion of the plank changes the normal flat surface of the

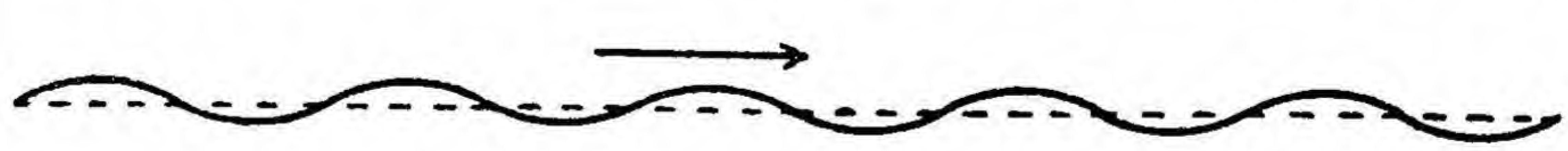


Fig. 175.

pond into a wavy shape consisting of a series of alternate crests and troughs, and it is this wavy shape which moves through the water (Fig. 175). There is little or no bodily transference of the water itself, as may be seen by putting a cork on the surface of the water, when it merely bobs up and down without moving in the direction of the waves. It is true that there is obviously transference of the water in waves formed near

the seashore, but these are a complicated kind of wave and we shall not consider them further. The above two kinds of wave are called **transverse waves**, because the vibration of any particle of the medium is at right angles to the direction in which the waves travel.

Then there are sound waves. When middle C is played on a violin, the string makes 256 vibrations every second and causes the belly of the violin to move up and down 256 times every second. When it moves up, it compresses the air just above it, and this compression travels out into the air; when it moves down, it rarefies the air above it, and this rarefaction follows the compression, 256 of each leaving the violin every second. This succession of compression and rarefactions is the condition which travels through the air in the case of sound, and it involves a vibration of the particles of air in the same direction as the waves travel, and so these waves are called **longitudinal waves**. We have seen that the passage of waves through a medium involves vibration of the particles of the medium, so we shall next consider a few simple properties of vibrations and then go on to make a model of waves.

109. SIMPLE PROPERTIES OF VIBRATIONS

A vibration is a to-and-fro motion which repeats itself at regular intervals. A good example is the motion of a pendulum or of a mass on a spiral spring (Fig. 176). The centre of the mass in each case moves to and fro along the path AOB, taking the same time to describe a complete cycle from A to B and back again. The **amplitude** of a vibration is the greatest displacement the body experiences measured from the equilibrium position. It is OA in the above cases. The **periodic time**, T, of a vibration is the time taken to execute one cycle or one to-and-fro motion. The **frequency**, f, of a vibration is the number of cycles executed in unit time. We have at once that

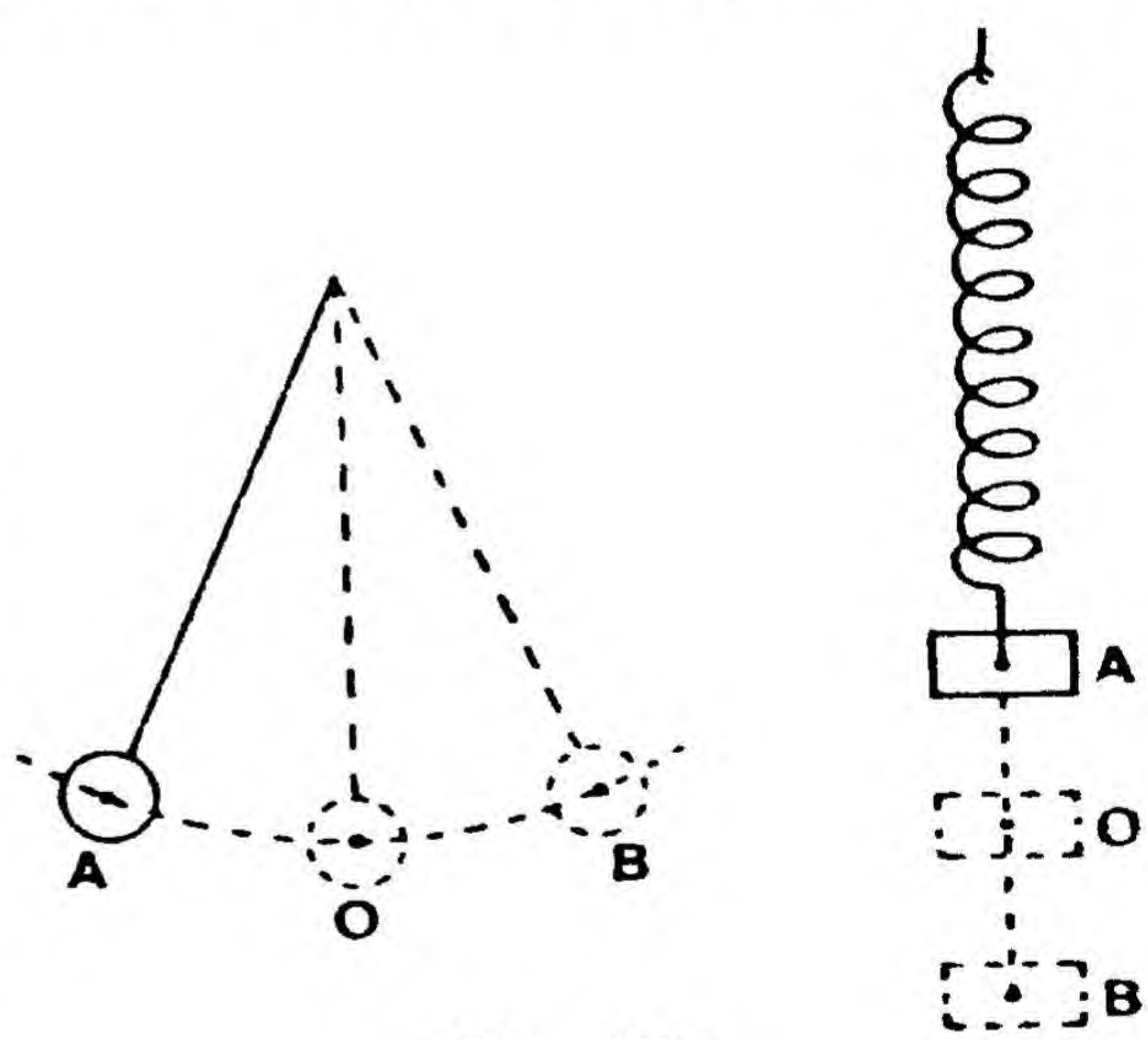


Fig. 176.

$$f=\frac{1}{T}$$

The simplest kind of vibration is Simple Harmonic Motion, in which the displacement y from the equilibrium position at any time t is given by

$$y=a \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

and the periodic time is independent of the amplitude. It will be discussed more fully in Art. 118.

110. THE IDEAL WAVE

As its name implies, this wave motion cannot be realised in practice, although water waves and sound waves of small amplitude come very close to it. It is hard to realise because of its simple properties. It consists of a wave form, which is represented by the graph of displacement of the particles against their position in the medium, travelling through the medium with type unchanged. This means that the shape of the wave form does not alter. To translate it into a practical case, it means that the shape of the ruffled surface of the water does not change as the waves move along. This is evidently not true for waves which break near the shore, but is nearly true for waves in deep water. Fig. 177

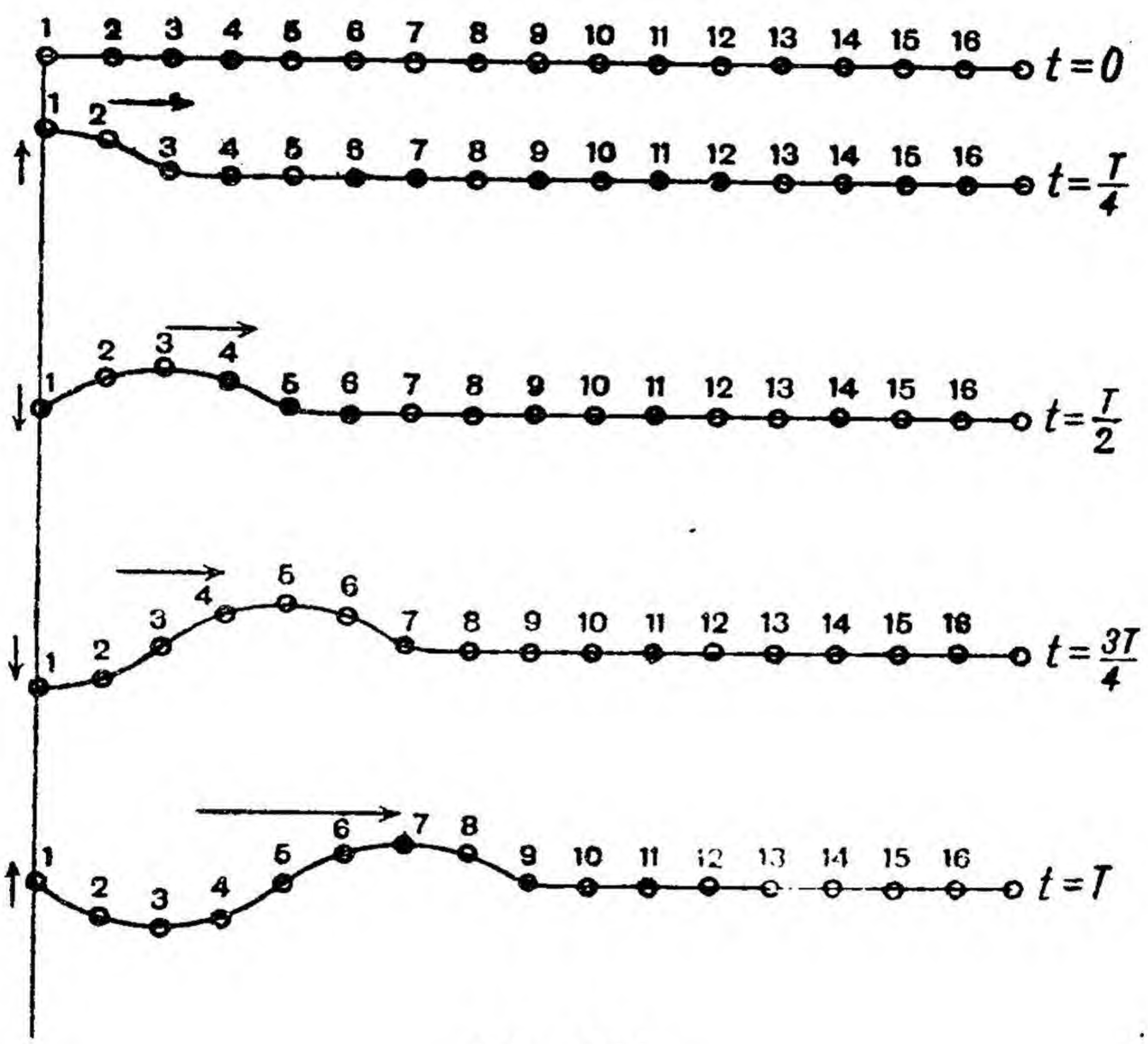


Fig. 177.

shows the wave being started by a vibration of particle 1 of the medium, which is handed on to the next particle and so on. This sets up a wave-form which travels on through the medium. The **wave-length**, λ, is the distance from one crest to the next, or the distance between two successive particles which are in identical states of motion (Fig. 178). The **amplitude**, *a*, of the wave is the distance from the crest to the normal position of the medium. The **period**, T, of the wave is the time taken by one complete wave to pass a given point in space, such as the undisplaced position of particle 5, and the **frequency**, *f*, of the wave is the number of complete waves passing a given point in space in unit time. **The velocity**, V, of the waves is the velocity of a given crest; if time is **reckoned** from the instant at which a given crest passes a given point

in space, then at the end of a second, f complete waves will have passed that point, and so the original crest will be a distance $f\lambda$ away. Therefore the velocity V of the waves is given by

$$V = f\lambda \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

We shall now suppose that the wave motion has been established and is being maintained by the vibration of the particle 1, and we shall consider in detail the passage of the wave form through the limited portion

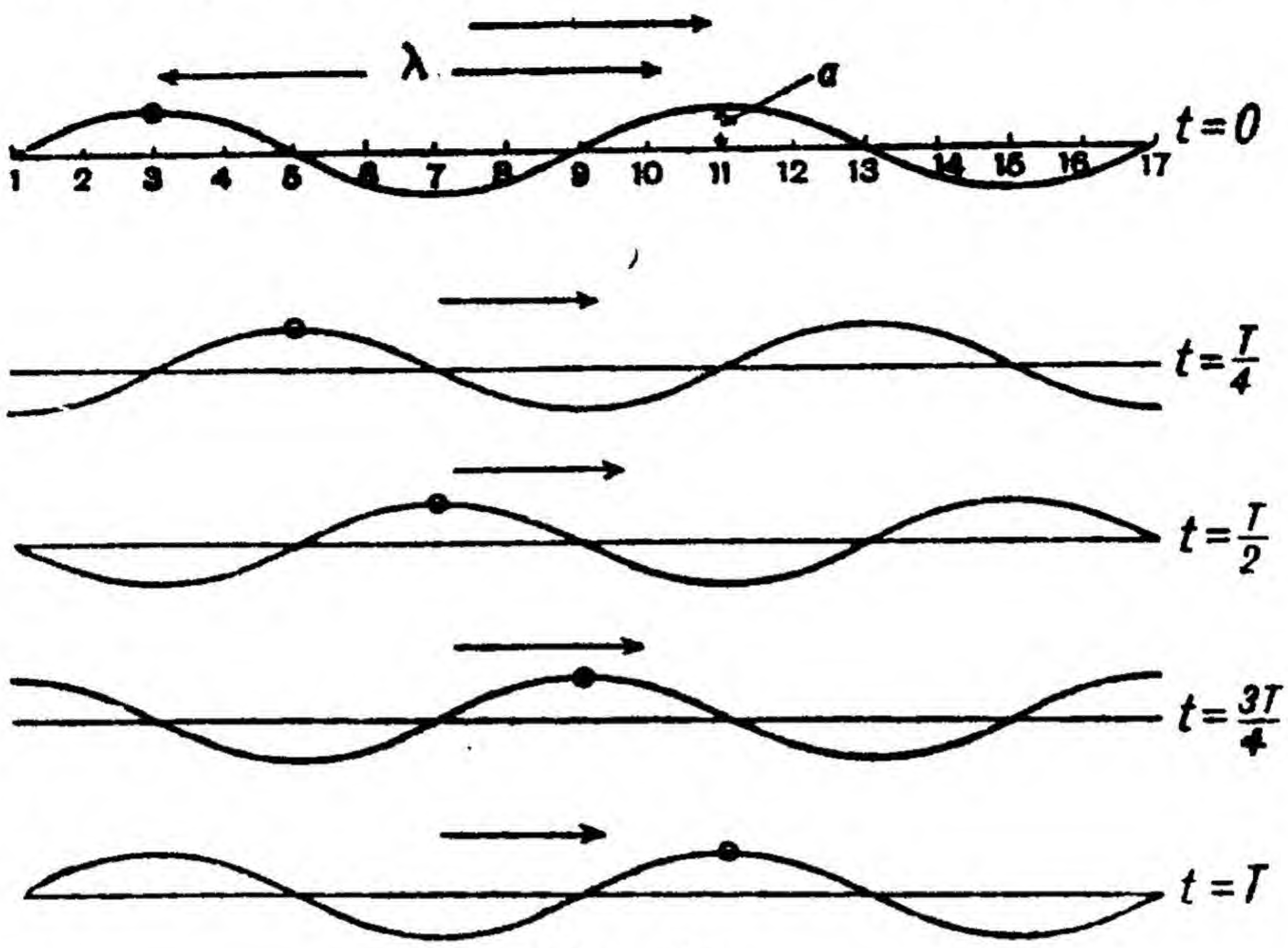


Fig. 178.

of the medium shown in Fig. 178. Let us consider the motion of particle 5. At $t=0$, it is in the position of zero displacement and is moving upwards; at $t=\frac{T}{4}$, it has reached the position of maximum displacement in the upwards direction; at $t=\frac{T}{2}$, it has reached the position of zero displacement once more, but it is moving in the downwards direction this time; at $t=\frac{3T}{4}$, it has reached the position of maximum displacement in the downwards direction, and at $t=T$, it is once more in the precise state in which it started, that is, at the position of zero displacement moving upwards. So while one complete wave passes the point in space normally occupied by the particle 5, that particle executes one complete vibration; in fact, the passage of the wave merely involves the vibration of the particle without any translation in the direction of the waves. If we now consider another particle, say particle 9, we see that the same thing is true in this case. It just executes vibrations as the wave passes along, one complete vibration occupying the same time as one complete wave takes to pass a given point in space. But the particles 5 and 9 do not vibrate in step, or in phase, to use the technical term: that is, they do not reach

the point of maximum upwards displacement simultaneously, nor do they attain any other corresponding point of their cycle of motion at the same time. As a matter of fact, these two particular particles are half a period or π out of phase, since 2π is taken to represent one period in the usual way of measuring phase. It is also true of any particle of the medium that the passage of the wave just involves a vibration of the particle whose period and amplitude are the same as those of the wave. **So the passage of any wave without change of type through a medium involves a vibration of the particles of the medium with the same period and amplitude but progressively varying phase.** For this reason the passage of such a wave does not produce any bodily motion of the medium in the direction of the waves, because the appropriate vibrations of the particles are all that is necessary to produce the motion of the wave form. And the amplitude, period, and frequency of the vibrations of the particles are equal to the corresponding quantities of the waves. We shall show later (Art. 119) that, if the wave form is a sine graph, the vibration of the particles of the medium is Simple Harmonic Motion.

What will happen if two separate waves pass through the same medium ? Suppose, for example, that two stones are dropped at the same time into a pond at a little distance from one another each producing circular waves. What will happen at the points where both waves meet ? If we consider one point in the medium, each wave is trying to set that particle into vibration with a certain period and amplitude. If the periods of the two waves are the same, then the resultant effect of the two waves at that point is to be found by adding together the two vibrations, paying due regard to the sign as well as the magnitude of the displacement due to each wave at any instant. Two particularly simple cases present themselves. In the first, the two vibrations due to the two waves are in phase. In this case, a crest from one wave arrives at the same time as a crest from the other and a trough with a trough. It is clear that the resultant vibration of the particle is one whose amplitude is the sum of the amplitudes of the component vibrations and the effect is a maximum in such a case. In the second case, the two vibrations due to the two waves are just half a period or π out of phase. This means that a trough from one wave arrives with a crest from the other or vice versa. So whatever effect one wave tries to produce on the particle, the other wave tries to produce precisely the opposite. So the resultant vibration is one whose amplitude is equal to the difference in the amplitudes of the component vibrations. It will be zero, if the amplitudes of the component vibrations are equal. The effect of the two waves is a minimum in this case. We shall meet with most important applications of this principle shortly (Art. 115).

111. HUYGENS' PRINCIPLE

So far we have considered waves in one direction only as along a rope or in one particular line on the surface of water. If a stone is dropped

into a pond, circular ripples spread out from the point O where the stone entered the water (Fig. 179). They are produced by the vibration of the water due to the entry of the stone and the ripples are circles with O as centre. Each one of these circles is called **a wave front** and the direction OA, in which any very small portion of a wave front travels, is normal to the wave front and is called a **ray**. It denotes the direction of propagation of the disturbance.

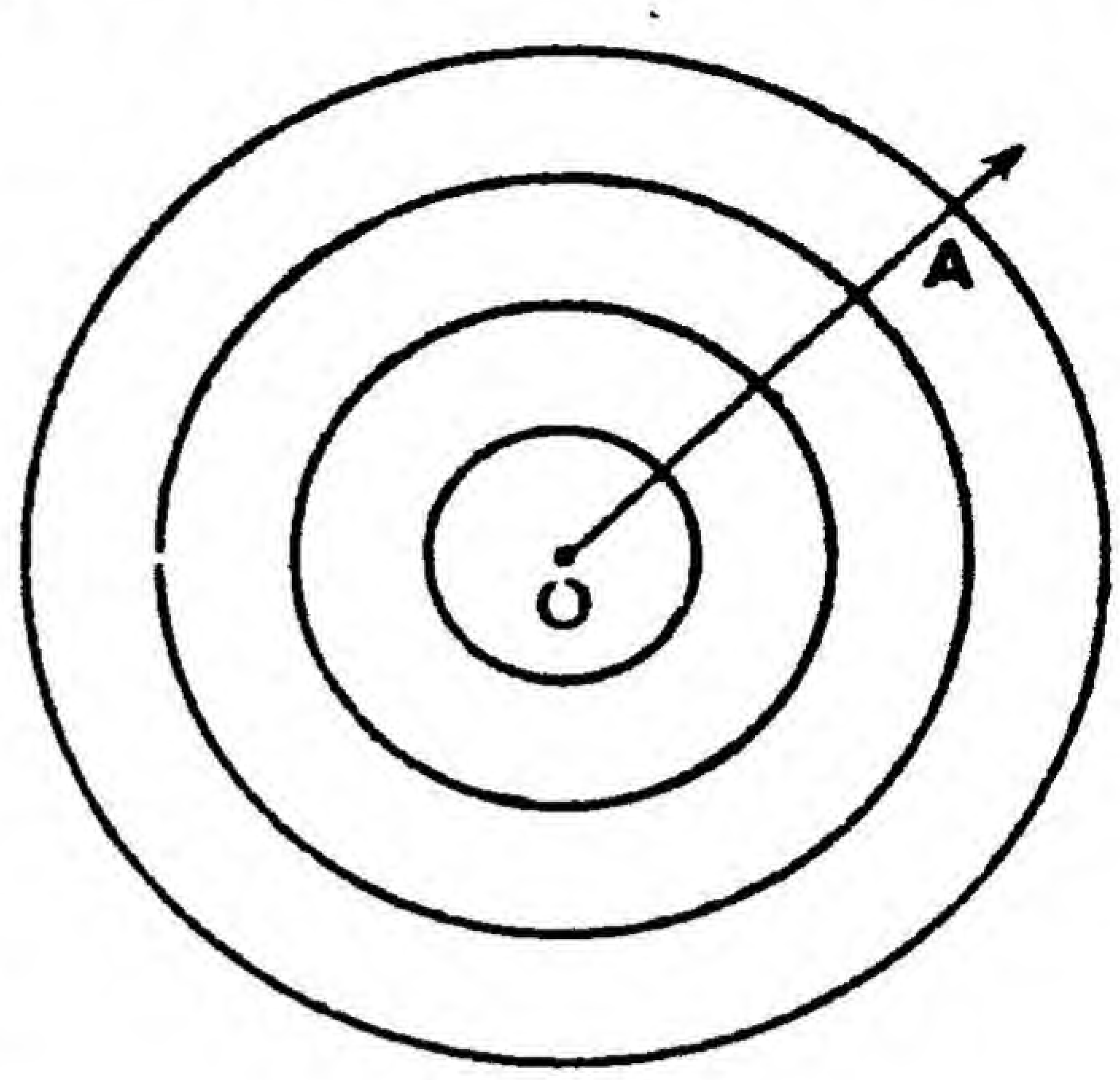


Fig. 179.

If a graph of the displacement of the particles, which normally lie on the line OA, against their distance from O is drawn, it will resemble the wave form shown in Fig. 175 and this wave form travels through the medium without any bodily motion of the medium itself. But we are now dealing with waves travelling in two dimensions and the conception of a wave front comes in. **A wave front is the locus of a particular phase of the vibration of the particles of the medium.** The phase most usually chosen is the position of maximum positive displacement, and so the locus is the line joining the crests. This is clearly a circle in the case of the waves produced by dropping a stone in a pond. If a plank AB floating on the surface of a pond is moved up and down, it will produce waves roughly of the form shown in Fig. 180, in which the wave fronts are straight lines, the rays being straight lines normal to the wave fronts. In the case of waves in three dimensions, a point source will produce spherical waves, the rays being radii of the spheres, while a plane source will produce plane waves in which the rays are lines normal to the planes.

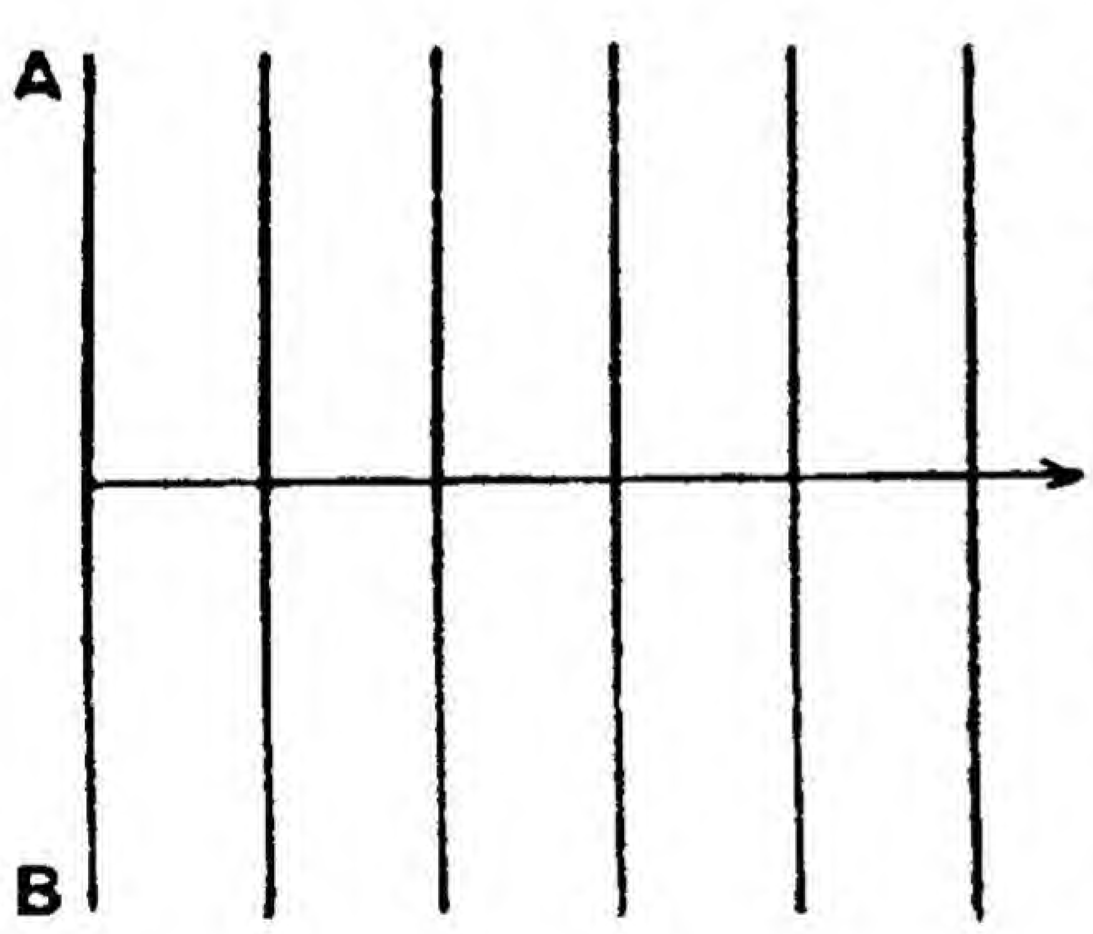


Fig. 180.

We know from simple experience that a circular wave front grows into a circle of ever increasing radius ; a linear wave front moves as a line parallel to itself ; a spherical wave front spreads as a sphere of ever increasing radius ; a plane wave front moves as a plane parallel to itself. What we need now is some theory to explain why this should be the case, some way of predicting the future course of wave fronts. This is essential if ever we are to put refraction and reflection on a rational basis. This theory was first put forward by Huygens and is known as **Huygens' Principle**. It states that the position of a wave front t seconds after its present position is found by regarding each point of the given wave front as the source of secondary wavelets. With each point on the given wave front as centre a sphere of radius ct is drawn, where c is the velocity of the waves in the medium. The required position of the wave front is the envelope of the

spherical wavelets. We shall illustrate this principle by applying it to the propagation of circular and plane waves.

Let us consider a circular wave front *abcde* . . . *qra* centre O (Fig. 181). If we wish to find the position of this wave front after a time t, we draw a set of spheres of radius ct with each point of the circular wave front as centre. A portion of the secondary wavelet from the points *abcd* . . . *qr* is shown in the diagram. The envelope, or common tangent, of these secondary wavelets is quite clearly another circle centre O and radius $Oa+ct$, thus verifying the fact that a circular wave front is propagated as a circle of ever increasing radius. Again, if we have a plane wave front

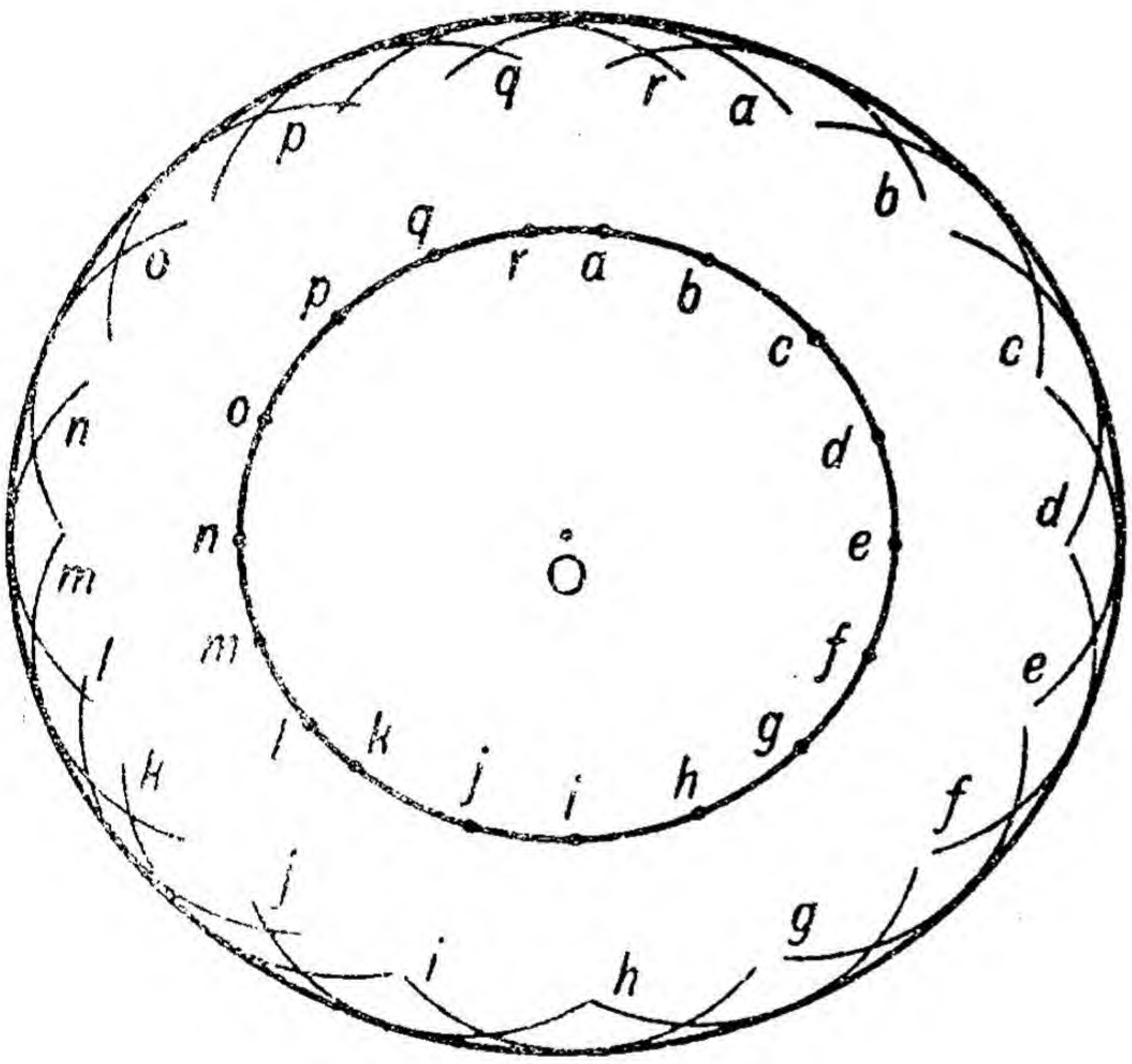


Fig. 181.

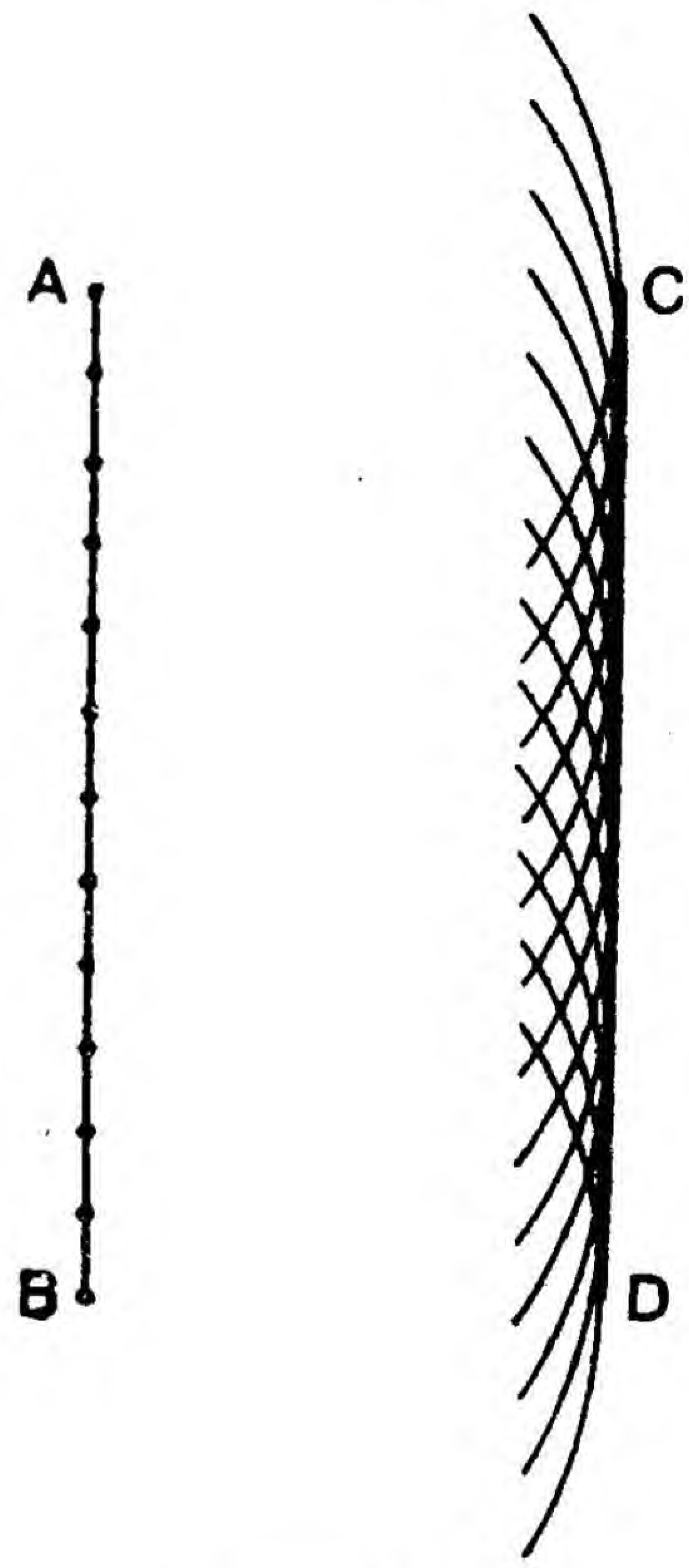


Fig. 182.

represented by AB, and we wish to find its position after a time t, we construct a set of spheres of radius ct with centres at each point on the plane wave front. A selection of these secondary wavelets is represented by the arcs of circles shown in Fig. 182 and the envelope of them is the line CD which represents a plane parallel to the original wave front and at a distance ct from it. This verifies the fact that a plane wave front is propagated as a plane, parallel to itself. But there are two difficulties to be noticed in connection with Huygens' principle. It is quite clear that in both the above cases we have only drawn *one* of the two possible envelopes to the secondary wavelets ; in the case of the circular wave front there is clearly one centre O and radius $Oa-ct$, and in the case of the plane wave front there is clearly one at the same distance from AB as CD but on the opposite side. So we have to know which way the wave is travelling before we can predict its future position. We have to know which envelope to ignore. **This is**

not a completely satisfactory position, but it need not hinder our using Huygens' principle as a working rule. We cannot accept it as the real explanation of how waves are propagated, nor can we regard the secondary wavelets as having real existence, until we can account for the absence of one envelope in a rational way. The second difficulty refers to the arcs of the secondary wavelets extending beyond CD in Fig. 182. What do they mean ? Do they indicate that a plane wave spreads out at its edges to a slight extent ? We say to a slight extent since the arc is just the effect of one wavelet, whereas the portion CD is the sum of many wavelets. It is known that waves do spread out sideways, so these arcs may indicate this diffraction. We shall refer to diffraction again later in this chapter, but we shall now proceed to the explanation of reflection and refraction on Huygens' principle.

112. REFLECTION

Let us suppose that a plane wave front, which is the same as a parallel beam of rays, is approaching a plane mirror PQ. We will use Huygens' principle to find the reflected wave front and see if it obeys the laws of reflection. Let AB (Fig. 183) represent the wave front at the instant when

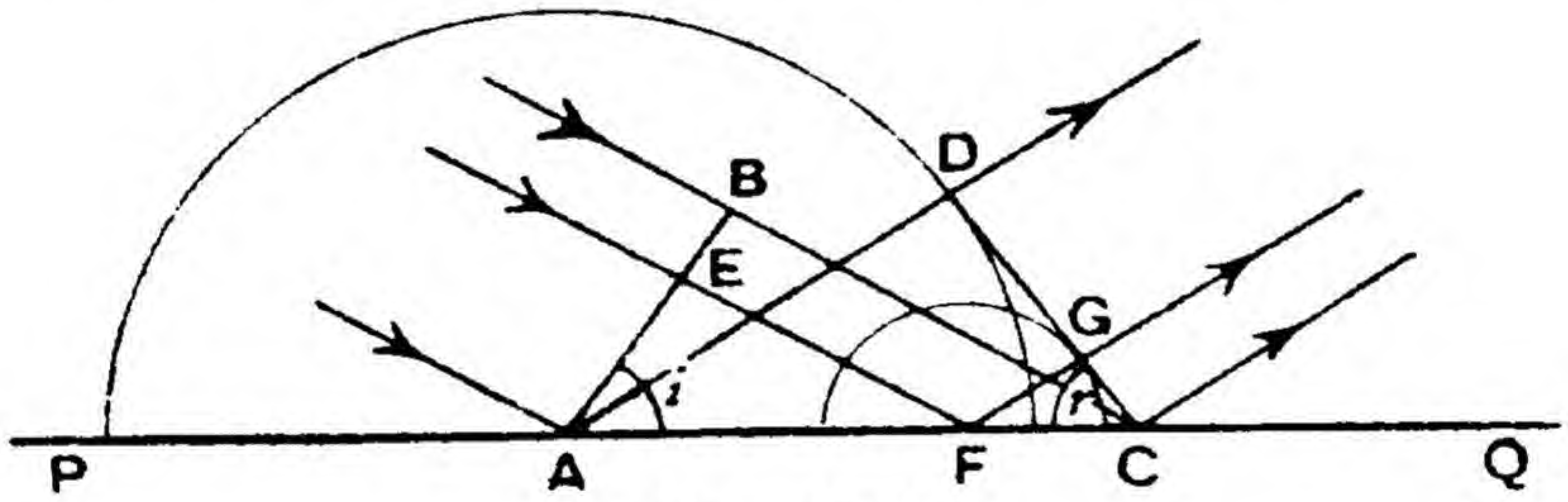


Fig. 183.

the end A has just reached the mirror. To find the position of the reflected wave front when the end B has just reached the mirror, we make use of the fact that the secondary wavelet starting from the point A will have a radius BC by the time the end B of the wave front has reached the mirror at C. So we draw a circle centre A and radius BC, drawing only the portion of the circle lying above the mirror, since we are seeking for the *reflected* wave front. The reader will recall that Huygens' principle cannot predict the way the wave front will travel. We then draw the tangent CD from C to this secondary wavelet. Now, in the right-angled triangles ABC and ADC, AC is common and BC=AD by construction, so the triangles are congruent and $\angle B\hat{A}C = \angle D\hat{C}A$. The truth of this relation is independent of the width AB of the incident wave front, and so if a smaller portion BE is considered, the tangent from C to the wavelet from F will make the same angle with the mirror as the tangent from C to the wavelet from A. In fact, the line CD touches all the wavelets from the points between A and C and represents the reflected plane wave front. We have proved that $\angle B\hat{A}C = \angle D\hat{C}A$. But $\angle B\hat{A}C$ between the incident

wave front and the reflecting surface is equal to the angle between the incident ray and normal to the reflecting surface, which is the angle of incidence, i; and $\angle D\hat{C}A$ between the reflected wave front and the reflecting surface is equal to that between the reflected ray and the normal, which is the angle of reflection, r. Thus $i=r$, which is the second law of reflection. The tangent from C to the spherical wavelet, centre A, radius BC, clearly touches the sphere at a point in the plane of the diagram and so the reflected ray is in the same plane as the incident ray and the normal to the reflecting surface at the point of incidence, which is the first law of reflection.

113. REFRACTION

Let AB represent a plane wave front approaching a plane surface separating two media in which the velocities of light are c_1 and c_2 respectively, the wave front proceeding from the less to the more dense medium (Fig. 184). To construct the refracted wave front at the instant when B

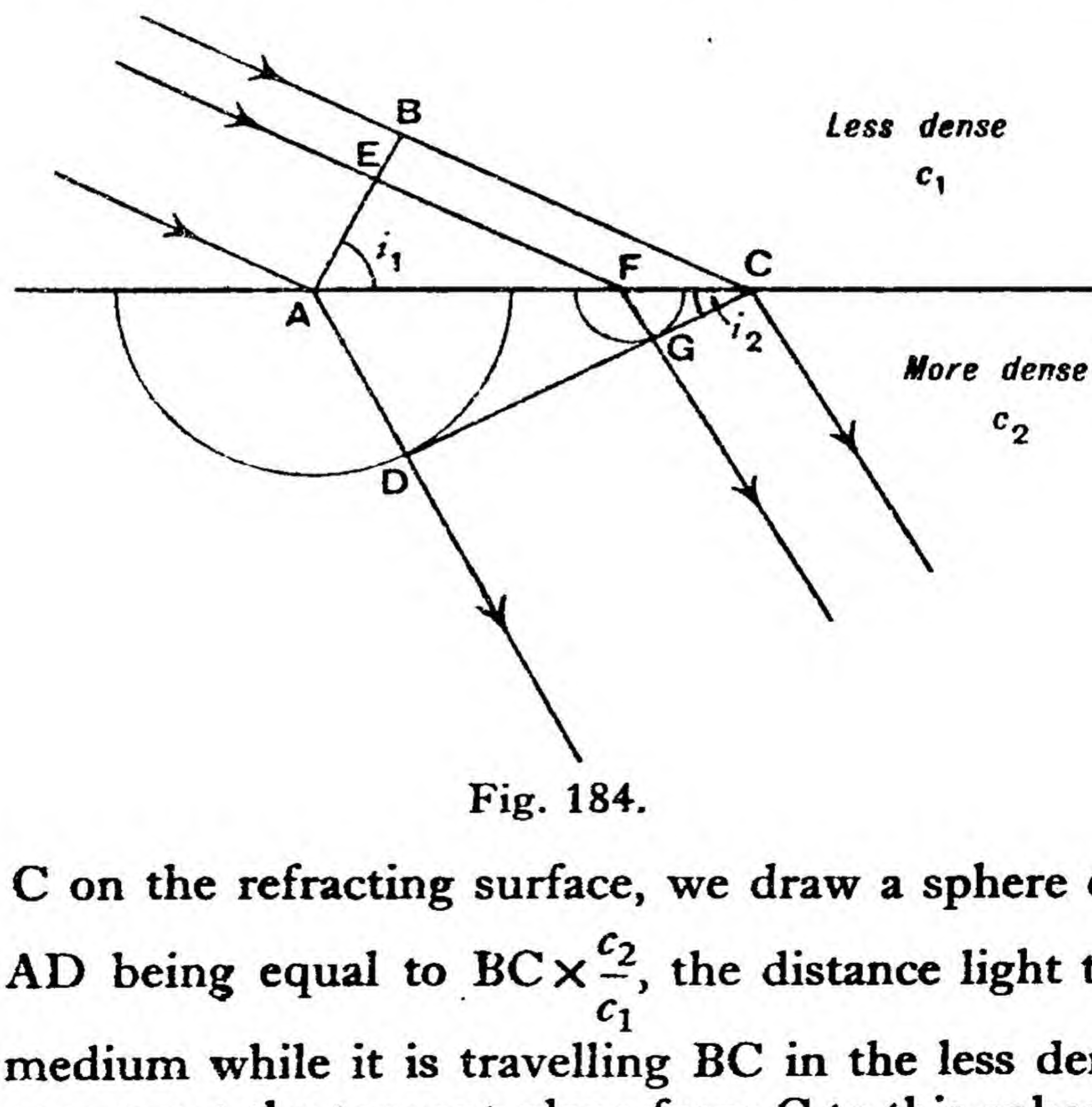


Fig. 184.

has reached C on the refracting surface, we draw a sphere centre A and radius AD, AD being equal to $BC \times \frac{c_2}{c_1}$, the distance light travels in the more dense medium while it is travelling BC in the less dense medium. Draw CD to represent the tangent plane from C to this spherical wavelet. We have

$$\frac{\sin i_2}{\sin i_1} = \frac{\frac{AD}{AC}}{\frac{BC}{AC}} = \frac{AD}{BC} = \frac{c_2}{c_1}$$

This gives a value of i_2 independent of the width AB of the wave front. If a smaller width BE is considered, CD will therefore be the tangent

plane to the secondary wavelet at F and it is therefore the tangent plane to the secondary wavelets from all the points between C and A. Accordingly CD represents the required refracted wave front. But i_1 and i_2 are the angles of inclination of the rays in the two media respectively and we have

$$\frac{\sin i_1}{\sin i_2}=\frac{c_1}{c_2}$$

a constant, which agrees with the general law of refraction

$$n_1 \sin i_1 = n_2 \sin i_2$$

This view of refraction leads to an interesting interpretation of the general law of refraction. The ratio of the sines of the angles of inclination for two media is constant, because it is equal to the ratio of the velocities of light in the two media. This ratio is in inverse proportion to the ratio of the refractive indices of the two media, thus showing that the velocity of light is smaller in the medium of greater refractive index, which has been shown to be true for air and water. Finally, the ratio of the velocities of light in air and water should be equal to the refractive index of water, and this has been verified experimentally (Art. 103), so the wave theory of refraction gives a complete explanation of the phenomena. It should therefore give an interpretation of the breakdown of refraction and its replacement by total internal reflection. The reader is advised to go into this for himself by answering question 4 at the end of this chapter.

114. THE LENS

Since the two axioms of geometrical optics, the laws of reflection and refraction, follow from Huygens' principle, the theorems about lenses and mirrors can be deduced directly from that principle too. We shall only

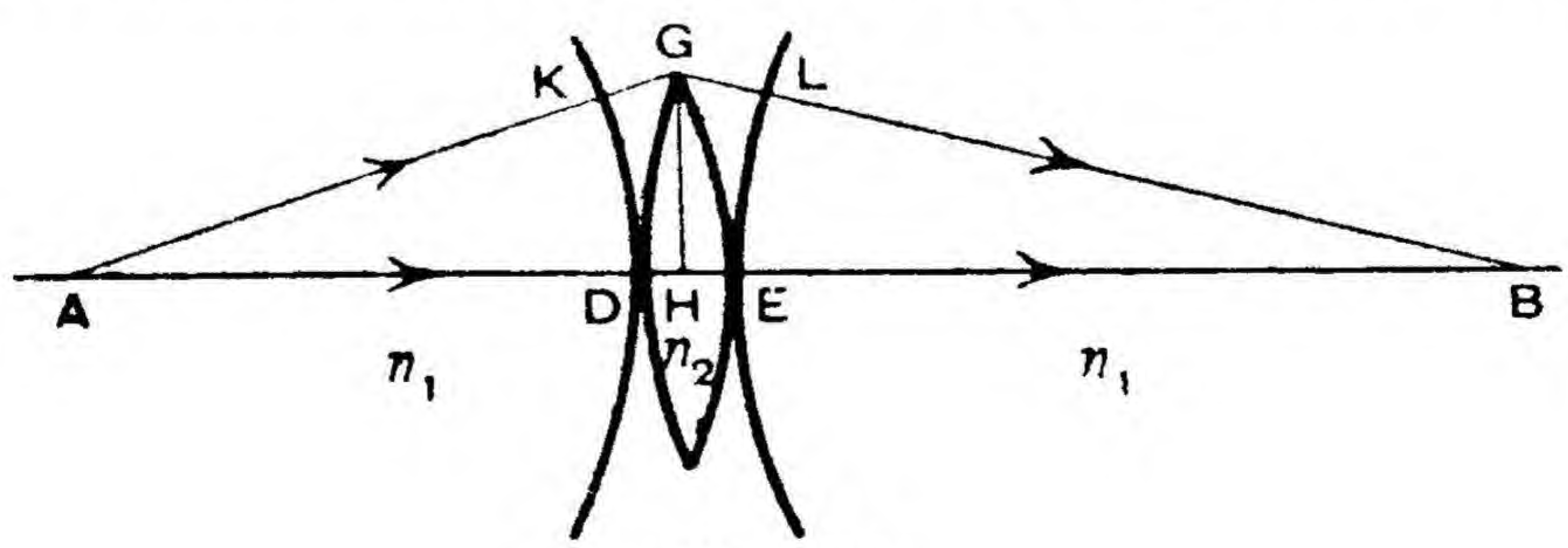


Fig. 185.

consider one of them, namely the case of a point image formed from an object on the axis of a converging lens. If A is a point object on the axis of a converging lens, it will produce a spherical wave represented by KD touching the surface of the lens at D (Fig. 185). Since light travels more slowly in the material of the lens than in the surrounding medium, the portion of the wave front round and about D will travel more slowly than the outer parts and the wave front will be turned inside out to become spherical and concave in the direction of motion. It is represented by

LE with centre B, which is therefore the point image of the point object A. Since the wave diverges from A and converges to B again, the time taken by light to travel the path ADHEB is the same as that taken to travel the path AKGLB. The reader will recall that this is the same principle which was used in Art. 19 in proving the lens theorem by Fermat's principle and so the proof of the lens formula from Huygens' principle follows just the same lines and will not be repeated here. Putting AH$=a$, BH$=b$, and the numerical values of the radii of curvature of the faces of the lens equal to c_1 and c_2, we obtain the usual relation

$$\frac{n_1}{b}+\frac{n_1}{a}=(n_2-n_1)\left(\frac{1}{c_1}+\frac{1}{c_2}\right)$$

and, if d is the numerical value of the focal length of the lens,

$$\frac{n_1}{d}=(n_2-n_1)\left(\frac{1}{c_1}+\frac{1}{c_2}\right)$$

and
$$\frac{1}{b}+\frac{1}{a}=\frac{1}{d}$$

There is an interesting interpretation of this last equation on the wave theory. $\frac{1}{a}$ is the curvature of the incident wave front when it strikes the lens and $\frac{1}{b}$ is the curvature of the emergent wave front on leaving the lens and $\frac{1}{d}$ is the curvature which the lens imposes on a plane wave front. So the curvature of the emergent wave front is equal to the curvature which the lens imposes on a plane wave front minus the curvature of the incident wave front. This only applies to the particular case of a converging lens producing a real image of a real object. The general case is obtained from the general equation

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

from which the curvature of the emergent wave front is equal to the sum of the curvatures of the incident wave front and that imposed by the lens on a plane wave front. The curvature of a wave front is reckoned positive or negative according as it is concave or convex when viewed by an eye placed so that light enters it after emerging from the lens. This view of a lens as an instrument which produces a constant change in curvature in a wave front falling on it is complementary to regarding it as a means of producing a constant deviation in rays falling on the lens at the same distance from the axis. In the one case, the power of the lens is the change in curvature which it produces in a wave front, while in the other it is the constant angle through which it deviates any ray falling on it at unit distance from the axis.

115. INTERFERENCE

We must now consider the phenomenon of interference which is quite well known in water waves, sound waves, and wireless waves. It refers to what is observed in the region where two or more wave trains cross. The word interference is perhaps unfortunate, since it implies that the individual wave trains interfere with one another in crossing, whereas, as a matter of fact, they emerge from the crossing bearing no trace of having been crossed by any other waves. But the word is too well established to admit of change. The reader must remember that no effects are to be observed except at the place where the crossing occurs. If a U-shaped piece of thick copper wire is mounted on the end of a vibrating strip of metal fixed to one side of a glass-bottomed tank filled to a suitable depth with water, each end of the copper can be made to dip into the

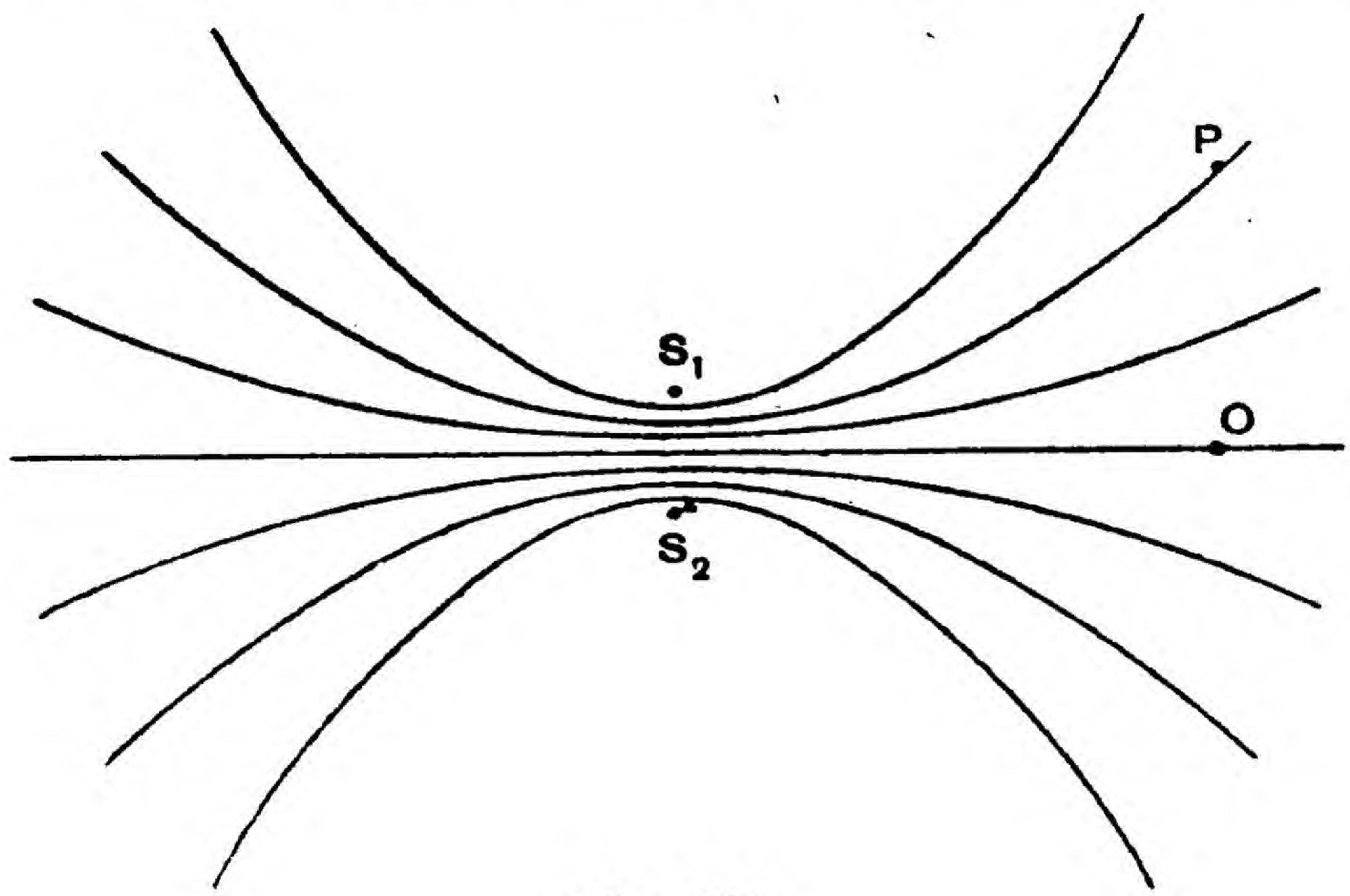


Fig. 186.

surface of the water and will act as a point source of water waves, if the strip is set vibrating. The waves produced on the surface of the water can be made visible to a class by sending light from an arc through the water and letting it fall either on the ceiling or reflecting it off a plane mirror on to a screen. The surface of the water is now disturbed by circular ripples from each point source and as these ripples are crossing each other at all points on the surface of the water, interference can be observed all over it. The appearance is shown in Plate IV and it can be seen that there are **a number of patches of calm water, separated by patches of water which are greatly disturbed by waves passing through them.** If the reader actually sees the phenomenon for himself he will also see that, although the waves are moving, the patches of alternate calm and disturbed water are stationary. They are called **interference fringes.** The explanation of the phenomenon will be clear from Fig. 186, which represents two point sources of waves, S_1 and S_2, sending

out spherical waves. Let us consider the effect at a point P. It will be a maximum or minimum according as S_2P-S_1P is $n\lambda$ or $(2n+1)\frac{\lambda}{2}$, where n is any integer and λ is the wave-length of the waves. For, if S_2P-S_1P is $n\lambda$, then a crest from S_1 will arrive at P at the same time as a crest from S_2 and their displacements of the particle of the medium at P will add up to produce a vibration of maximum amplitude. Also the vibrations of the particle of the medium at P due to the waves from S_1 and S_2 will always be exactly in step and so P will remain permanently a place of maximum disturbance, provided S_2P-S_1P is $n\lambda$. On the other hand, if S_2P-S_1P is $(2n+1)\frac{\lambda}{2}$, then a crest from S_2 will arrive at P at the same time as a trough from S_1 and their displacements of the particle of the medium at P will be in opposite directions and if their amplitudes are equal they will cancel out. So the displacement of the particle at P will be zero and it will remain zero, since the displacements of the particle at P due to the waves from S_1 and S_2 will always be equal and opposite. The locus of an interference fringe is found by putting n constant and is therefore given by

$$S_2P-S_1P=\text{constant}$$

This is a set of hyperboloids of revolution about S_1S_2 as axis and the effect observed with water waves is the intersection of the surface of the water with these hyperboloids of revolution, there being one for each value of n. This explains the alternate stationary patches of calm and disturbed water and it can be seen that the patches of calm in Plate IV are approximately hyperbolæ. The striking characteristic of this phenomenon is that *disturbance added to disturbance produces calm.* But this does not involve any loss of energy, since the energy apparently lost in the regions of calm is present in the extra energy existing in the regions of disturbance. It is important to notice three characteristics of these interference fringes produced by two point sources. The fringe width, or distance between the centres of two patches of calm, increases as we go further from the sources, as is obvious from Plate IV; the fringe width increases as the distance between the sources decreases, as is shown by Plate IV, Figs. 1 and 2; and it decreases as the wave-length decreases. This is illustrated in Plate IV, Figs. 1 and 3, but it can also be seen by considering the effect at a point such as O on the central fringe of zero path difference (Fig. 186). If a line be drawn through O perpendicular to the central fringe, the next fringe of maximum disturbance will occur when a point so far from O along that line is reached that the path difference is one wave-length. If the wave-length is decreased, it is not necessary to go so far to one side of O before the necessary path difference is reached and so the fringe width gets less. *Any attempt to produce interference with two point sources of light must produce interference fringes showing the above characteristic features.*

PLATE IV

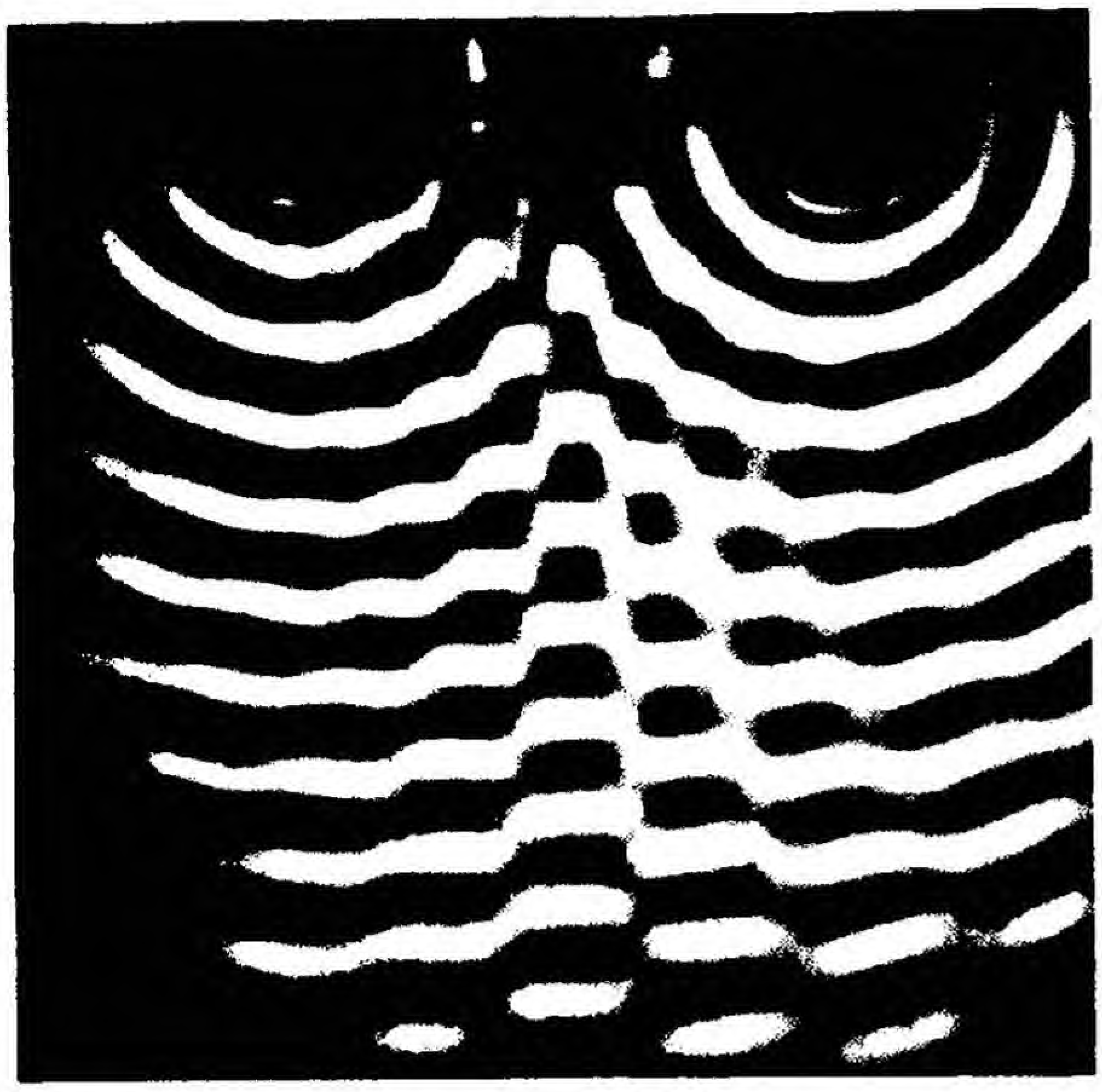


Fig. 1. Interference fringes in water waves with two point sources wide apart, producing a small fringe width

(*J. W. Cottingham*)

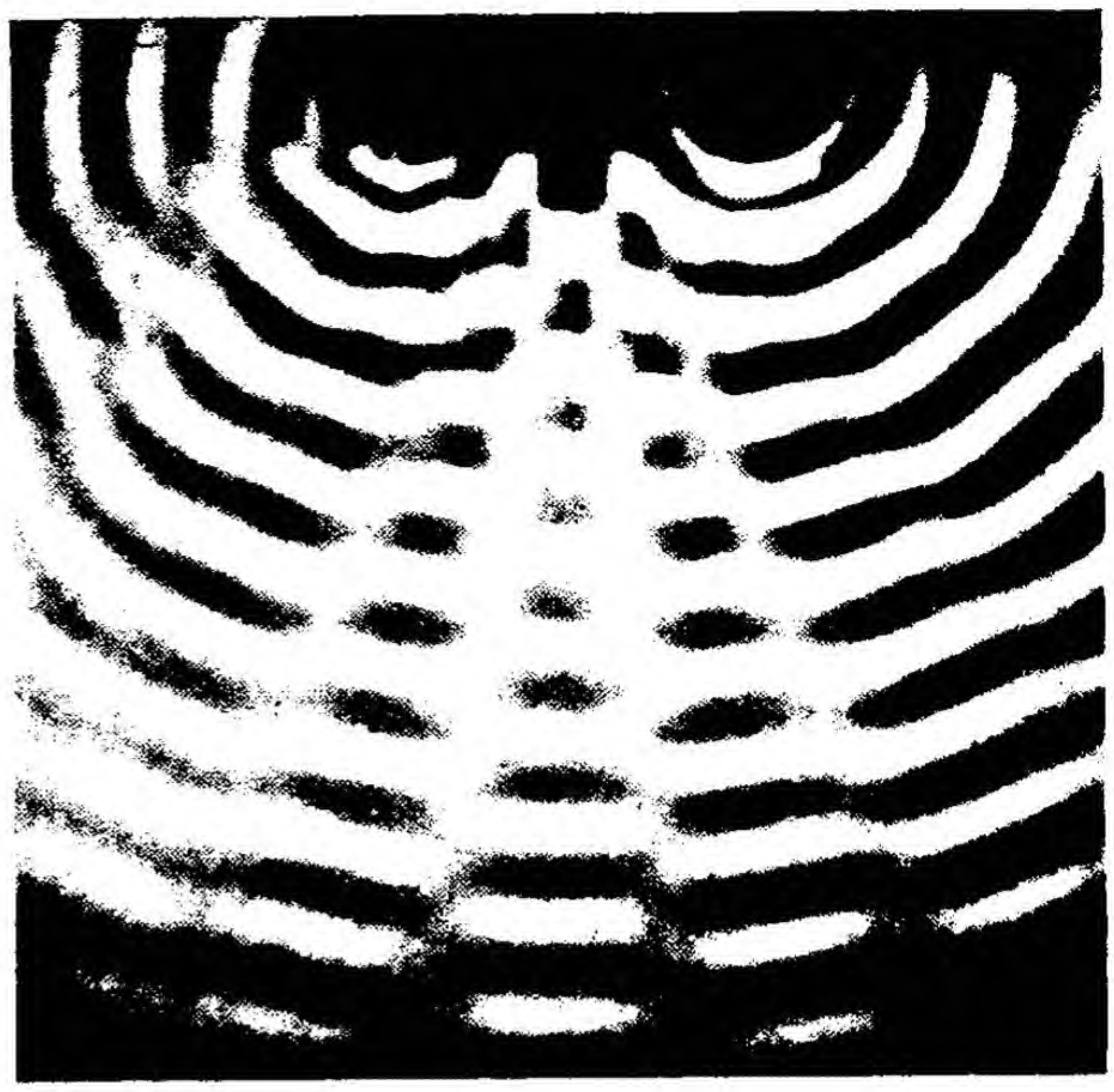


Fig. 2. Interference fringes in water waves with two point sources close together, producing a large fringe width.

(*J. W. Cottingham*)

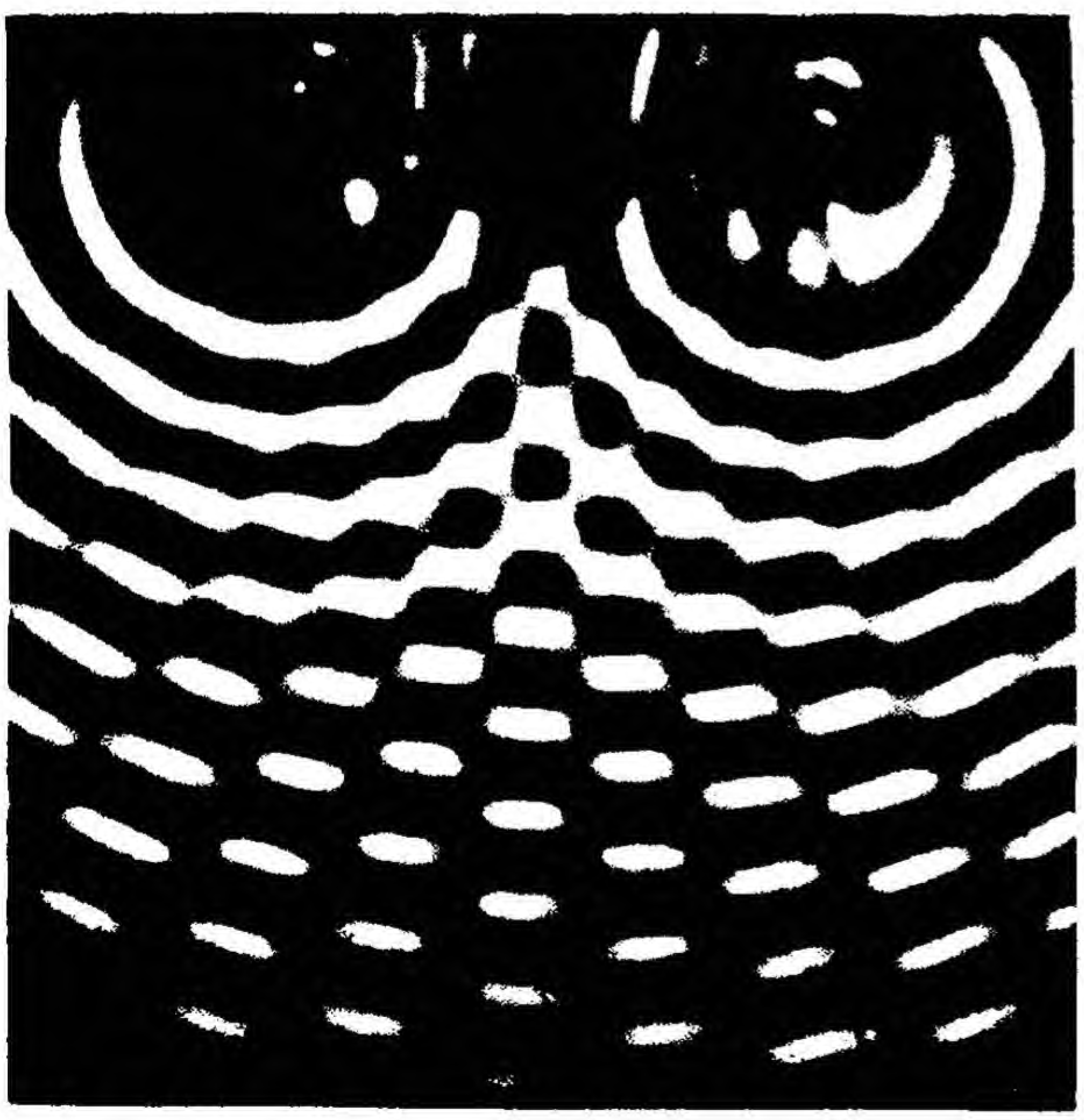


Fig. 3. Interference fringes in water waves with two point sources the same distance apart as in Fig. 1, but with a smaller wave-length, thus producing a smaller fringe width.

(*J. W. Cottingham*)

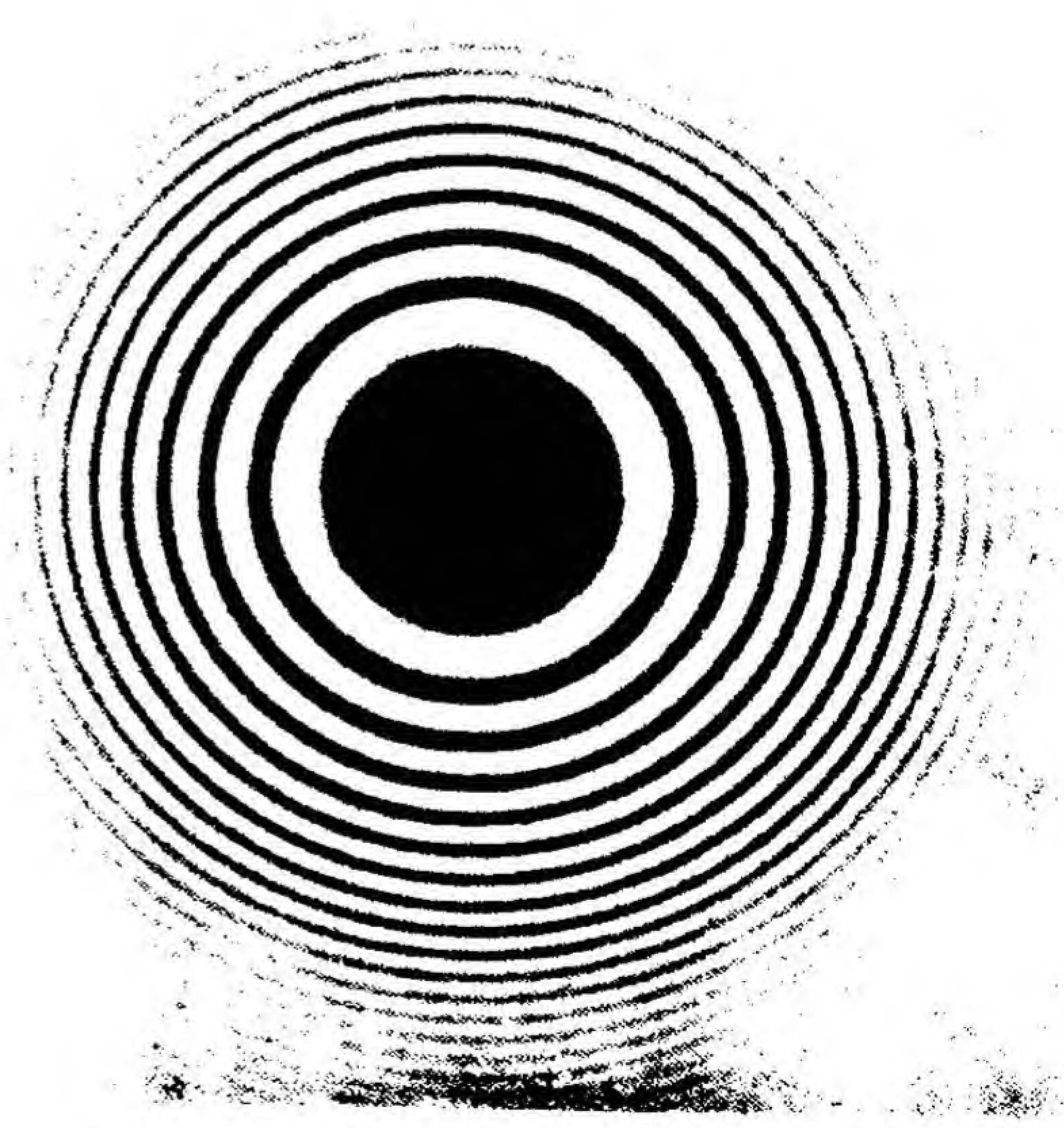


Fig. 4. Newton's Rings.

(*J. W. Cottingham*)

There are many examples of interference in sound and we will consider two simple cases. If a tuning-fork is struck and held close to the ear with the prongs vertical and is rotated about a vertical axis, no sound will be heard four times in each revolution. The explanation is as follows. When the prongs GG (Fig. 187) move outwards, a rarefaction is produced in between them and moves out along the lines OY and OY′, while at the same time compressions are produced on the outside of the prongs and are propagated along the lines OX and OX′. But the compressions and rarefactions spread out over spherical wave fronts and so there will be regions of silence along the lines POP′ and QOQ′, the bisectors of YOX′ and YOX respectively, since a compression from outside GG and a rarefaction from in between GG will arrive at any point of these lines simultaneously. Therefore the four silences heard in each revolution of the fork are due to the fringes of silence OP, OQ, OP′, and OQ′ sweeping over the ear in succession. A second case is the phenomenon of

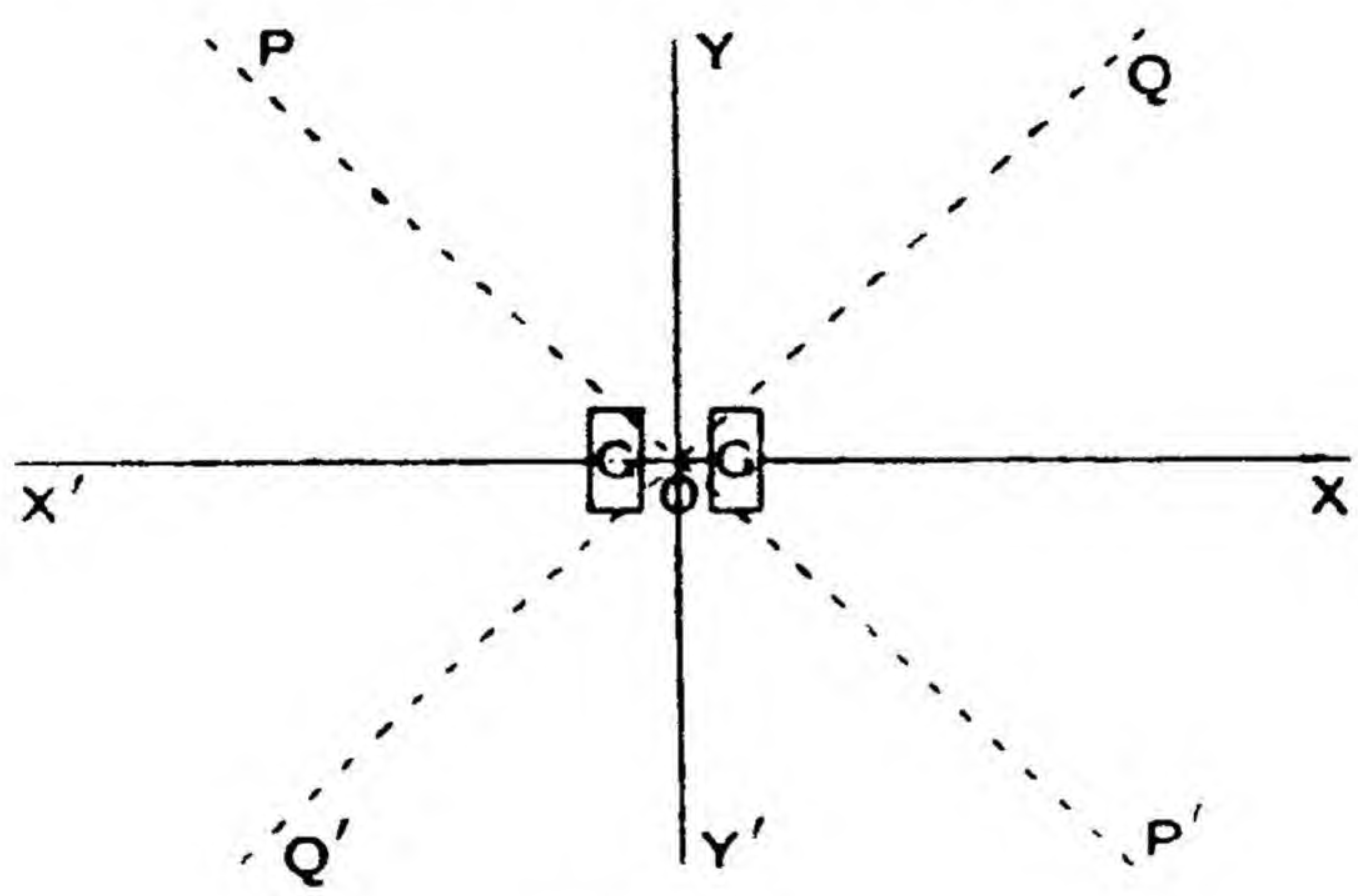


Fig. 187.

beats, which is the periodic waxing and waning in intensity of the sound heard when two tuning forks of nearly the same frequency are vibrating at the same time. Let S_1 and S_2 be the two forks (Fig. 186) and let us imagine that the observer is at O equidistant from each fork. Let the two forks have frequencies of 256 and 258 c.p.s. respectively. Reckoning time from the instant when each fork is sending out a compression, the effect at O is a maximum, since the compressions arrive there at the same time and their effects add up. But a quarter of a second later one fork has completed 64 vibrations and the other 64·5 vibrations, so the former is sending out a compression at the same time that the latter emits a rarefaction. Therefore the effect at O is a minimum, or, if the amplitudes of the two forks are the same, silence. In half a second the forks have executed 128 and 129 vibrations respectively, so that each is sending out a compression and the effect is a maximum, or loudness, once more. This effect continues regularly as is summarised in Table 15, and the loudness of the sound waxes and wanes at 2 c.p.s., the difference of the component frequencies, which agrees with the observations.

TABLE 15

Time sec.	Number of Vibrations Executed by		Nature of Disturbance Emitted by		Effect at O
	S_1	S_2	S_1	S_2	
0	0	0	Compression.	Compression.	Maximum.
$\frac{1}{4}$	64	64·5	Compression.	Rarefaction.	Minimum.
$\frac{1}{2}$	128	129	Compression.	Compression.	Maximum.
$\frac{3}{4}$	192	193·5	Compression.	Rarefaction.	Minimum.
1	256	258	Compression.	Compression.	Maximum.

Here again we may sum up the essential feature of interference as **sound added to sound produces silence** and again this paradox is not a violation of the conservation of energy, as the absence of sound in the patches of silence is compensated by excess of sound in the patches of loudness.

Finally we may refer to a case of interference in the electric waves, which are used to transmit music and speech in broadcasting. The condition which is propagated in this case is an electric and magnetic field, which are at right angles both to each other and the direction of the waves, and the wave form shown in Fig. 175 represents an electric wave if the ordinates represent the strength of the electric field in space. It is known that electric waves can travel right round the earth, because there is a region of the atmosphere some 30 miles up called the ionosphere, which reflects electric waves down to the earth again. When electric waves do travel right round the earth, it is because they have suffered several reflections both at the ionosphere and the surface of the earth itself. Accordingly, when a broadcasting programme is received on a wireless set, some of the incoming waves may have travelled directly along the surface of the earth and some may have gone up to the ionosphere and been reflected down again. It is possible for interference to occur between these two sets of waves, and if a crest of the direct wave happens to arrive at the receiver at the same time as a trough of the reflected wave, the resultant will be a minimum and the reception will be poor. Fortunately the height of the layer in the ionosphere from which the waves are reflected downwards is continually altering and the interference fringe of minimum effect may move away from the receiving set to another place and be replaced by a fringe of maximum effect. Later the minimum may return and cause the reception to become poor or to fade, using the technical term, and this fading is one of the difficulties which modern wireless reception is trying to overcome by arranging for the set to amplify more when the amplitude of the incoming wave decreases. But, for us, the interest in fading is that it is yet another example of interference, and once again we see that **electric field added to electric field produces zero field.**

This is the feature common to all the cases of interference we have so far considered, that disturbance added to disturbance produces calm and it is difficult to see how such a phenomenon could be produced by anything but waves with their crests and troughs. Conversely, if interference is obtained, it is fair to argue that we are dealing with waves.

116. DIFFRACTION

The last phenomenon characteristic of water waves and sound waves is diffraction. The diffraction of water waves is illustrated in Plate V, in which plane waves are incident on an aperture which is some five times wider than the wave-length of the waves, and it is seen that, after passing through the aperture, the waves spread sideways behind the obstacles which form the aperture. There is no question of a sharp shadow being cast as is the case with light; the disturbance definitely extends into the geometrical shadow, and the rays, which are perpendicular to the wave fronts, bend round instead of travelling in straight lines. Further, if the width of the aperture is decreased, the extent of the bending or diffraction increases in the sense that the amplitude of the waves inside the geometrical shadow increases and also that the waves spread further into the shadow itself. Finally, if the aperture is about the same width as the wave-length of the waves, it acts practically as a point source and the waves spread out equally in all directions from it as is shown in Plate V. Precisely similar effects are obtained in the case of obstacles; waves bend round to a small extent behind an obstacle some five times wider than their wave-length and the intensity and extent of this bending increase as the width of the obstacle decreases. These effects can be produced in a ripple-tank, but they also occur naturally on the sea or a lake. Even the most casual observer must have noticed that the ripples of small wave-length produced by a faint breeze in the pool of water around a large boulder at low tide do not spread into the region behind the boulder, while the large waves flowing towards the same boulder at high tide pass by and close in again after a few yards bearing hardly any trace of their temporary break into two parts by the boulder. Diffraction, then, is **the sideways spreading of waves after they have passed through apertures or by obstacles and becomes more marked as the width of the obstacle or aperture decreases so as to become the same as that of the wave-length of the waves themselves.**

The same effect occurs with sound waves. If we stand on one platform of a big railway station and listen to a train, which is blowing off steam from the safety valve, coming in at another platform, we can hear the low-pitched clank of the train quite clearly although a block of buildings may be directly between the train and us. This is because the clank of the train produces waves of wave-length about 10 ft. and so they can spread sideways behind buildings some 10 to 20 ft. long and reach us. But the hiss of the steam from the safety valve is a high-pitched sound and the wave-

length of the waves it emits is only an inch or two, so that these waves cannot spread sideways behind the large buildings to any extent. Therefore we do not hear the hiss of the steam from the safety valve until we can see the engine ; that is, until the waves can come directly from their source to us. Everyone must have noticed this effect and it is another example of the fact that waves spread sideways behind obstacles and that this sideways spreading is the more marked as the wave-length of the waves decreases to become of the same length as the size of the obstacle. We shall not go into the theory of diffraction here, beyond remarking that it must be based on Huygens' principle and that the propagation of plane waves on that principle led us to expect a sideways spreading of plane waves.

117. PREDICTIONS

The way is now clearly signposted for the supporters of the wave theory and the signposts are labelled **interference** and **diffraction.** We must realise interference in the case of light ; *our task is to add light to light and produce darkness* ! Secondly, we must explain why light is propagated so nearly in straight lines (the reader will remember that the experimental evidence of Art. 2 does not prove that light travels exactly in straight lines) and we must produce diffraction in light. We must set up an experiment which shows quite clearly light rays bending round the corners of an obstacle or spreading out after passing through an aperture. Are these quests doomed to failure at the outset ? Are they hopeless ? Is it not possible that light shows so little diffraction because its wave-length is so small ? And what is the explanation of the dark bands outside the geometrical shadow in Grimaldi's experiment (Art. 97) and the alternate dark and bright bands inside the shadow ? Are the bright bands inside the shadow a case of diffraction ? Possibly this experiment contains the germ of the phenomena we are seeking. Let us press on then into the unknown and settle if light is or is not waves ! Just before starting we must consider briefly the mathematical theory of waves, which we shall need at a later stage.

118. SIMPLE HARMONIC MOTION

We have already seen that the passage of a wave through a medium involves vibration of the particles of the medium and so we naturally begin the mathematical theory of waves with a consideration of the simplest vibration, Simple Harmonic Motion.

Simple Harmonic Motion is the resolved part parallel to any direction of uniform circular motion. If a particle P (Fig. 188) describes a circle centre O and radius a with uniform *speed* v and uniform angular velocity ω and a line PQ is drawn perpendicular to a diameter YOY′ of the circle to cut the diameter at Q, then the point Q describes

simple harmonic motion. It is clear that Q describes a to-and-fro motion repeating itself at regular intervals, because, as P describes the circle starting at X and proceeding through Y, X′, Y′ back to X, Q goes from O to Y back to O on to Y′ and finally back to O again, having executed one cycle or complete vibration. We shall now derive expressions for the **displacement** y of the point Q measured from the point O, its **velocity** $\dot{y}$ and its **acceleration** $\ddot{y}$ at any time t, time being reckoned from the instant at which P is at X and Q is at O moving upwards.

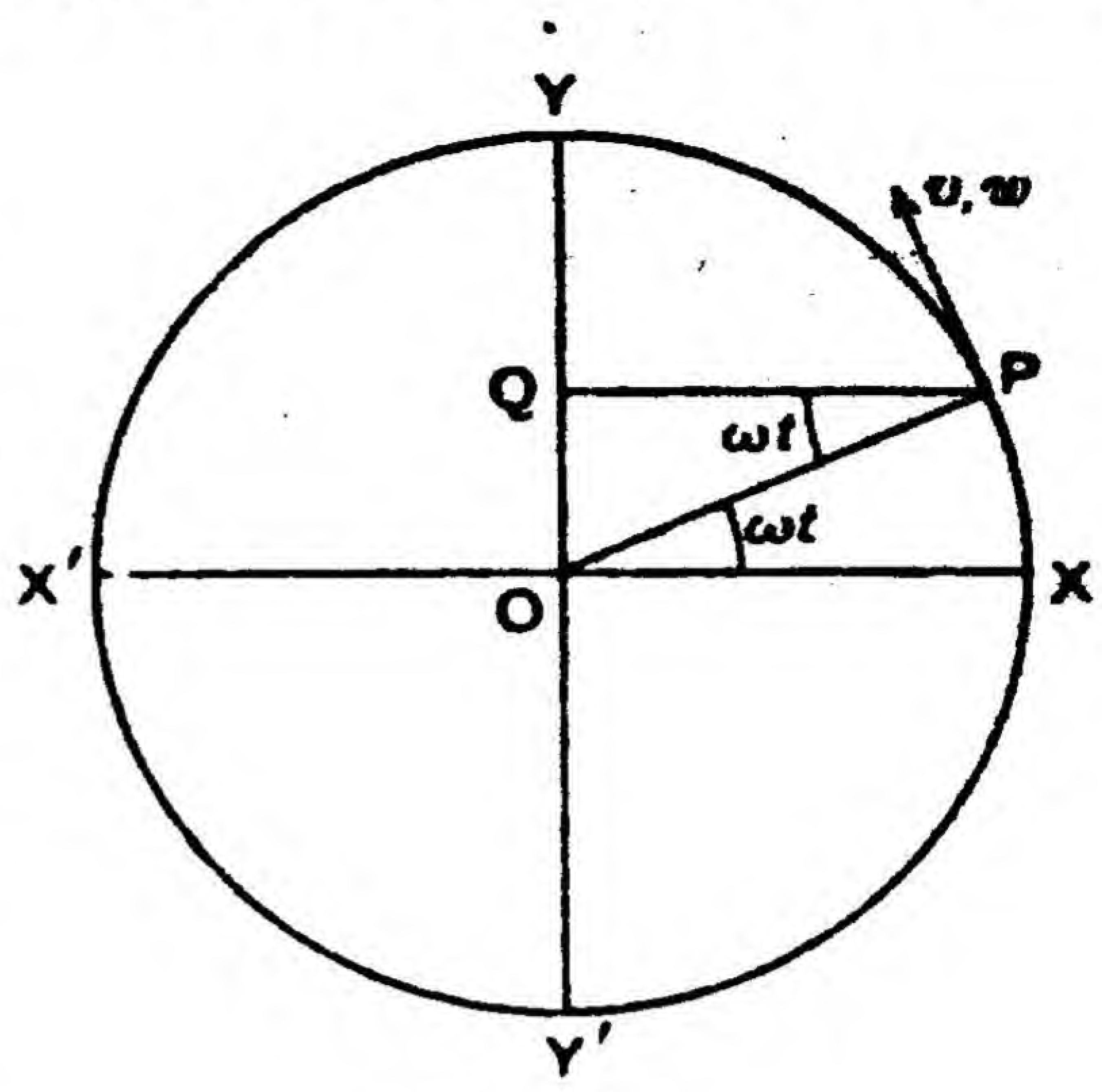


Fig. 188.

The angular velocity of a point describing a circle about any axis is the angle measured in radians described in unit time by the line joining the point to the axis. So the angular velocity of the point P is related to its speed in the circle by the equation

$$\omega = \frac{v}{a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)$$

If the point P reaches the position shown at time t, it has taken t seconds to travel from X to P and so the angle XOP is ωt. Therefore

$$OQ = OP \sin \omega t$$

or

$$y = a \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

The velocity of Q is the resolved part of the speed v of P in the direction YOY′. Since OP makes an angle ωt with XOX′, the direction of the velocity of P, being at right angles to OP, makes the same angle with the line YOY′ at right angles to XOX′. Therefore

$$\dot{y} = v \cos \omega t$$

From equation (60)

$$v = \omega a$$

$$\dot{y} = \omega a \cos \omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

a result which can also be obtained by differentiating equation (61) with respect to t.

The acceleration of Q is the resolved part of the acceleration of P parallel to YOY′. The acceleration of P is $\frac{v^2}{a}$, or $\omega^2 a$ from equation (60), along PO. Therefore

$$\ddot{y} = -\omega^2 a \cos\left(\frac{\pi}{2} - \omega t\right)$$

$$\therefore \ddot{y} = -\omega^2 a \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

Combining this equation with equation (61) we have the important result

$$\ddot{y}=-\omega^2 y \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (64)$$

The negative sign appears in this expression for the following reason. If Q is above O, the velocity of Q in the direction OY is always decreasing; in other words, if y is a positive number, $\ddot{y}$ is a negative number and the negative sign is inserted in the equation to ensure this result. If Q is below O, its velocity is always increasing; the reader must remember that a velocity of -6 say half-way between O and Y′ is less than the velocity O at Y′ itself. Mathematically, if y is a negative number, $\ddot{y}$ is a positive number and again the negative sign ensures that the equation shall lead to this result. In fact, the negative sign is the *mathematical* way of putting into the equation the *physical* fact that the acceleration is always in the *opposite* direction to the displacement. This last equation leads to another way of defining simple harmonic motion as **the motion of a particle in a straight line in which the acceleration of the particle is proportional to its displacement from the equilibrium position and oppositely directed.**

The **periodic** time T of the motion is the time taken for Q to describe one complete to-and-fro motion, which is the same as the time taken for P to go once round the circle. Therefore

$$T=\frac{2\pi}{w} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

This period is independent of the amplitude of the motion. Combining equations (61) and (65) we have for the equation of simple harmonic motion

$$y=a \sin \frac{2\pi}{T} t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

The **phase** of a simple harmonic motion at a given instant is the value of the angle POX or of the quantity ωt at that instant. It really denotes what stage of the cycle the particle has reached at that instant. If the phase is 0, the particle is at the position of zero displacement and moving towards Y, if it is $\frac{\pi}{2}$ the particle is at the position of maximum positive displacement, if it is π, it is at the position of zero displacement and moving towards Y′ and so on. The **phase difference** between two simple harmonic motions is an important quantity and is the difference in phase of the two motions at any instant. It will, of course, remain constant if the motions have the same periodic time. If the phase difference between two motions is π, one will be at the position of maximum positive while the other is at the position of maximum negative displacement; one will be at the point of zero displacement moving in one direction, while the

other is at the same place moving in the opposite direction. If the phase difference is $\frac{\pi}{2}$, one is at the position of zero displacement and moving towards Y, while the other is at Y; the one is always a quarter of a cycle behind the other. *The phase difference is how much the two vibrations are out of step with each other*; if it is 0, they are in step; if it is π, they are exactly out of step, if it is $\frac{\pi}{2}$, they are half out of step and so on. The reader should make sure that he understands this idea, since we shall be continually using it in later work.

119. THE EQUATION OF WAVE MOTION

We shall now derive the equation of wave motion. The wave form is a sine-graph shown in Fig. 189 of wave-length λ, amplitude a: if the

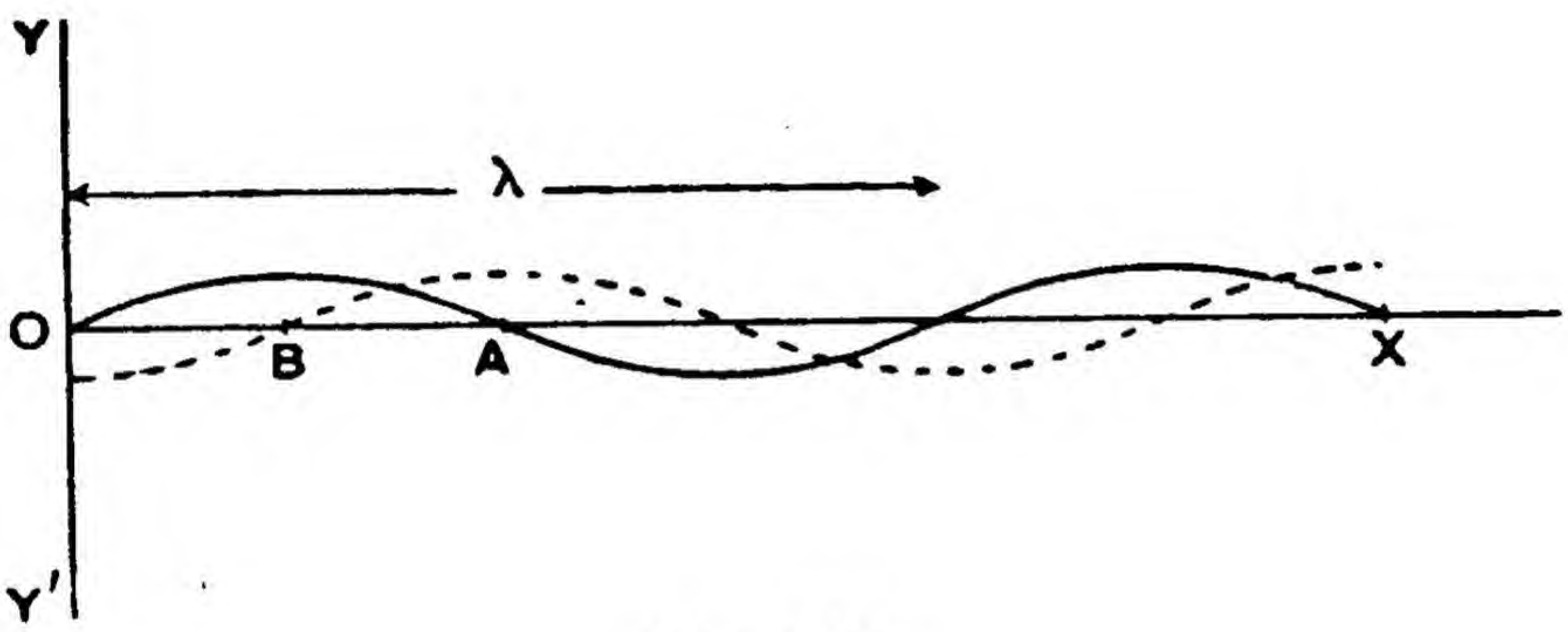


Fig. 189.

period of the waves is T, it is moving to the right with velocity $\frac{\lambda}{T}$. The diagram shows its position in the medium at time $t=0$ and the equation of the wave form is

$$y=a \sin \frac{2\pi}{\lambda} x \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (67)$$

which represents a sine-graph repeating itself at points separated by a distance λ, since an increase of x by the amount λ reproduces the same value of y. The dotted line in Fig. 189 shows the position of the wave at time $t=\frac{T}{4}$ and from these two curves it is clear that the displacement y of the particle A of the medium at time $t=\frac{T}{4}$ is the same as that of the particle B at $t=0$, where B is the same distance to the left of A $\left(\frac{\lambda}{4}\right)$ as the

wave travels in the time $\frac{T}{4}$. In general, then, the displacement y at time t of the particle at a distance x from O is the same as the displacement at $t=0$ of the particle at distance $\left(x-\frac{\lambda t}{T}\right)$ from O, since the wave moves a distance $\frac{\lambda t}{T}$ to the right in time t. The displacement at $t=0$ of the particle at distance $\left(x-\frac{\lambda t}{T}\right)$ from O is from equation (67) $a \sin \frac{2\pi}{\lambda}\left(x-\frac{\lambda t}{T}\right)$, which is also the displacement y at time t of the particle at distance x from O.

$$\therefore y=a \sin \frac{2\pi}{\lambda}\left(x-\frac{\lambda t}{T}\right)$$

$$\therefore y=a \sin 2\pi\left(\frac{x}{\lambda}-\frac{t}{T}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

which is the required general equation of a sine wave, wave-length λ, amplitude a, moving to the right with velocity $\frac{\lambda}{T}$. The reader will notice that y depends both on x and t, because the particles of the medium are displaced due to the imposition of a wave form on the medium and the wave form itself is moving.

The motion of any particular particle of the medium can be deduced by keeping x constant. If we consider the particle at O, putting $x=0$ in equation (68)

$$y=-a \sin \frac{2\pi}{T} t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (69)$$

which is the equation of a simple harmonic motion of period T. Thus the passage of the wave causes the particle to execute a S.H.M. of the same period and amplitude as the wave. Considering a particle at a constant distance x from O, its equation of motion is given by equation (68), x being constant, and this is the equation of S.H.M. of period T and amplitude a, the sole effect of the constant term in x being to introduce a phase difference of $2\pi \frac{x}{\lambda}$ between this particle and the particle at O. As x increases from 0 to λ, this phase difference increases from 0 to 2π, so that a particle at distance λ from O executes an S.H.M. exactly in phase with that of the particle at O. So the passage of the sine wave through the medium causes **each particle to describe a S.H.M. of the same period and amplitude as those of the wave but with progressively varying phase.** If the particles vibrate at right angles to the direction of propagation of the wave form, the waves are transverse; if they vibrate in the same direction, the waves are longitudinal.

We can also see that the equation represents the movement of the wave form by considering the relation between y and x at two instants, say $t=0$ and $t=t_1$. At the first instant the wave form is given by

$$y=a \sin 2\pi\frac{x}{\lambda}$$

and at the second instant by

$$y=a \sin 2\pi\left(\frac{x}{\lambda}-\frac{t_1}{T}\right)$$

Each equation, regarding t_1 as constant, represents a sine graph of wave-length λ; in the first case the least positive value of x for which $y=0$ is $x=0$. In the second case, the least positive value of x for which $y=0$ is $x=\frac{\lambda}{T}t_1$. So the wave form has travelled a distance $\frac{\lambda}{T}t_1$ to the right in time t_1 and therefore its velocity v is given by

$$v=\frac{\lambda}{T}=f\lambda. \quad \quad (70)$$

where f is the frequency of the waves.

All these deductions about the behaviour of this ideally simple wave agree with those deduced by the graphical method of Art. 110, but the mathematical analysis is necessary to show that *sine* waves involve *S.H.M.* of the particles of the medium. The graphical method can only prove that the propagation of a wave form with type unchanged involves just some vibration of the particles of the medium; it cannot give the connection between the type of wave form and the type of vibration.

120. THE COMPOSITION OF SIMPLE HARMONIC MOTIONS AT RIGHT ANGLES

As we shall come across examples of a particle executing two S.H.M.'s of the same period at the same time in directions at right angles in some of the phenomena of light, we will deduce the resultant of such motions here, so that we can use the result when we want it. Let the S.H.M.'s be along the axes of co-ordinates XOX′ and YOY′ (Fig. 190); if the displacements of the particle are x and y respectively at time t, then its resultant position is at P, whose co-ordinates are x and y. Our problem is to find the locus of P, which should be independent of t. We shall consider three simple cases which often occur in Light.

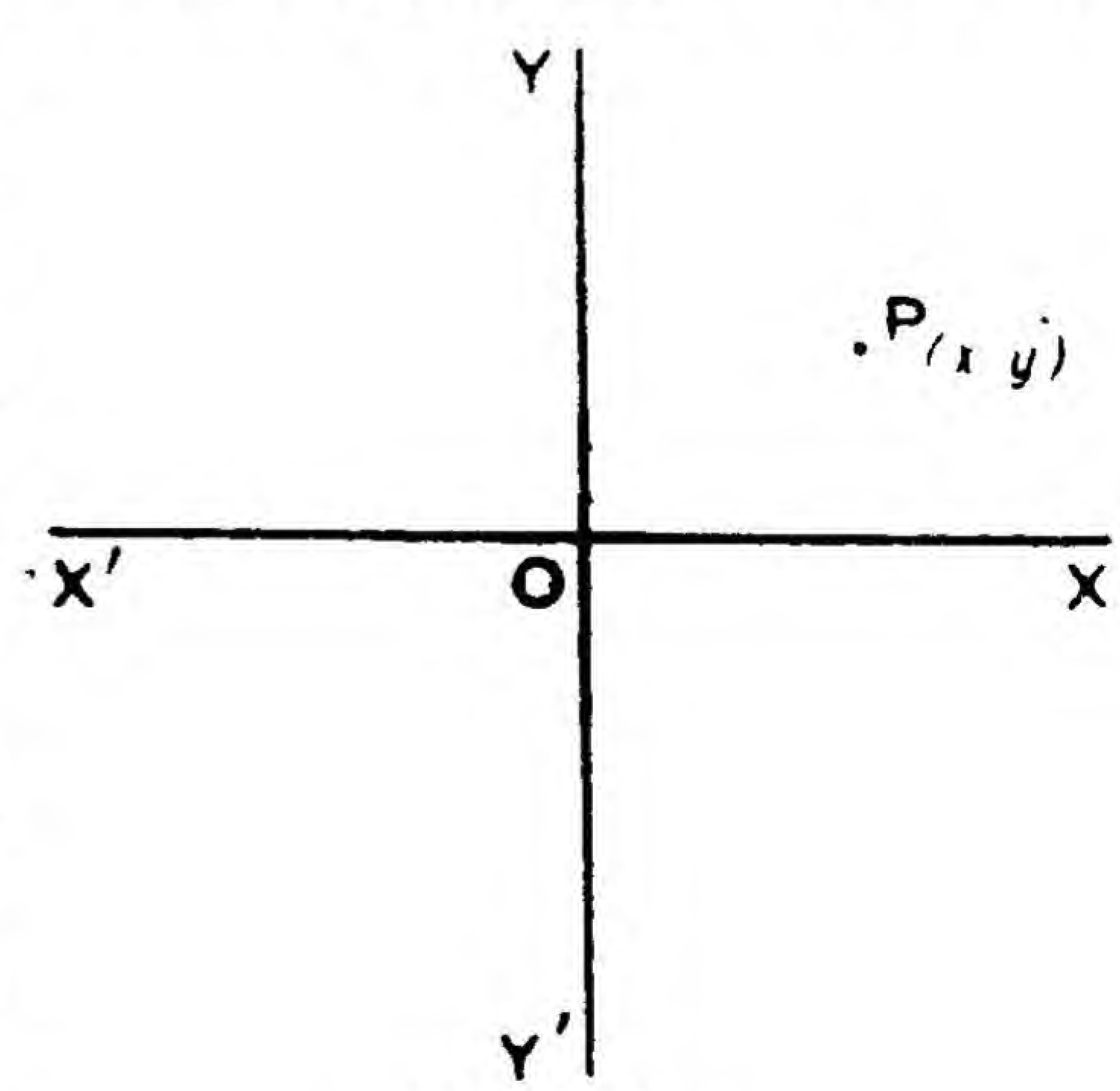


Fig. 190.

Case 1 : The S.H.M.'s are in phase and of different amplitude.
Let them be represented by the equations

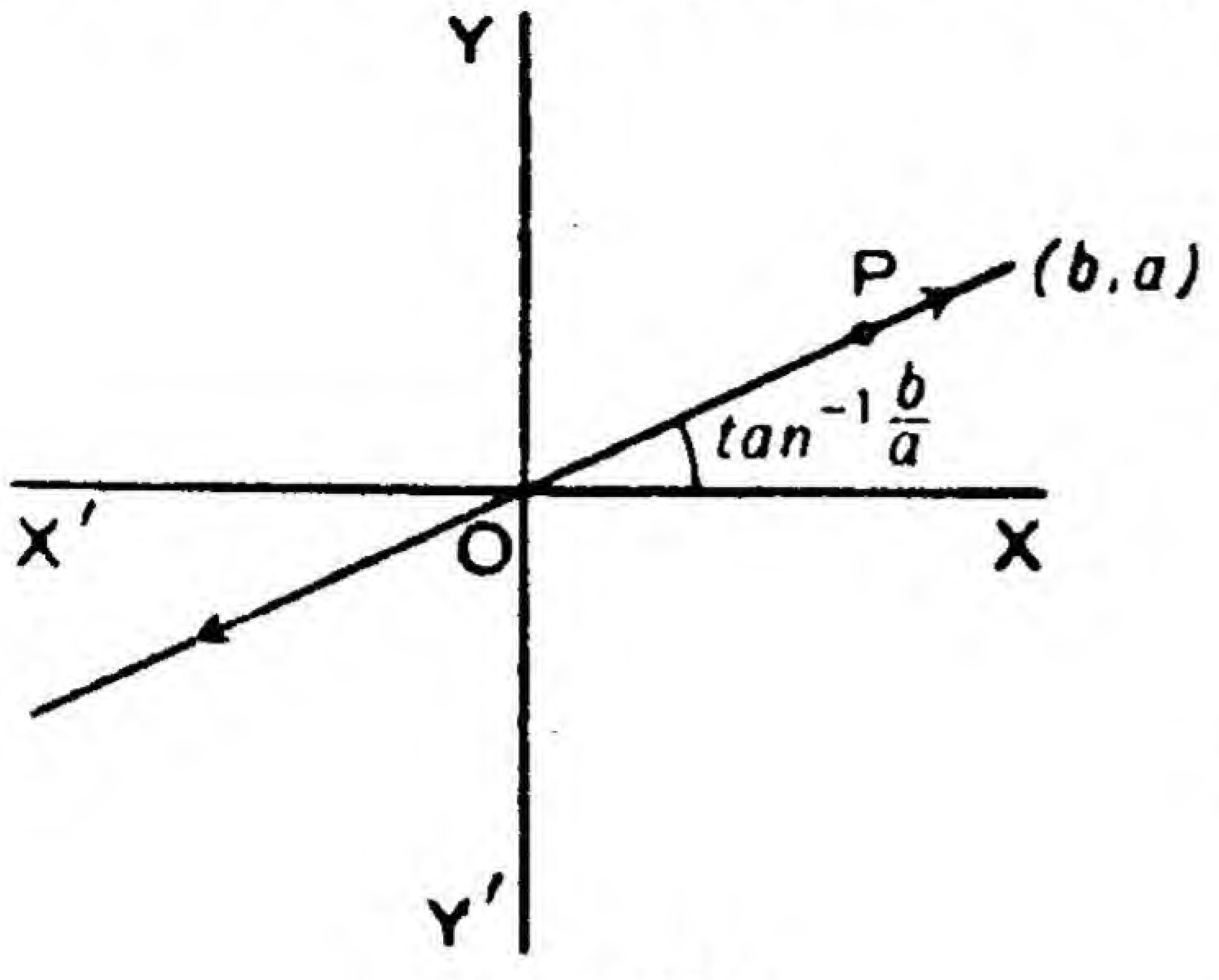


Fig. 191.

$$x = a \sin \omega t$$

$$y = b \sin \omega t$$

Dividing the two equations we have

$$\frac{y}{x} = \frac{b}{a}$$

which is the equation to a straight line making an angle $\tan^{-1} \frac{b}{a}$ with the x axis (Fig. 191). The particle describes a to-and-fro motion along this line in the same time as the common period of the S.H.M.'s. If the amplitudes are equal, the line is at 45° to the axes of co-ordinates.

Case 2 : The S.H.M.'s are $\frac{\pi}{2}$ out of phase and the amplitudes are unequal.

Let them be represented by the equations

$$x = a \sin \omega t$$

$$y = b \sin \left(\omega t + \frac{\pi}{2}\right)$$

$$\text{or } y = b \cos \omega t$$

$$\therefore \frac{x^2}{a^2} = \sin^2 \omega t$$

and

$$\frac{y^2}{b^2} = \cos^2 \omega t$$

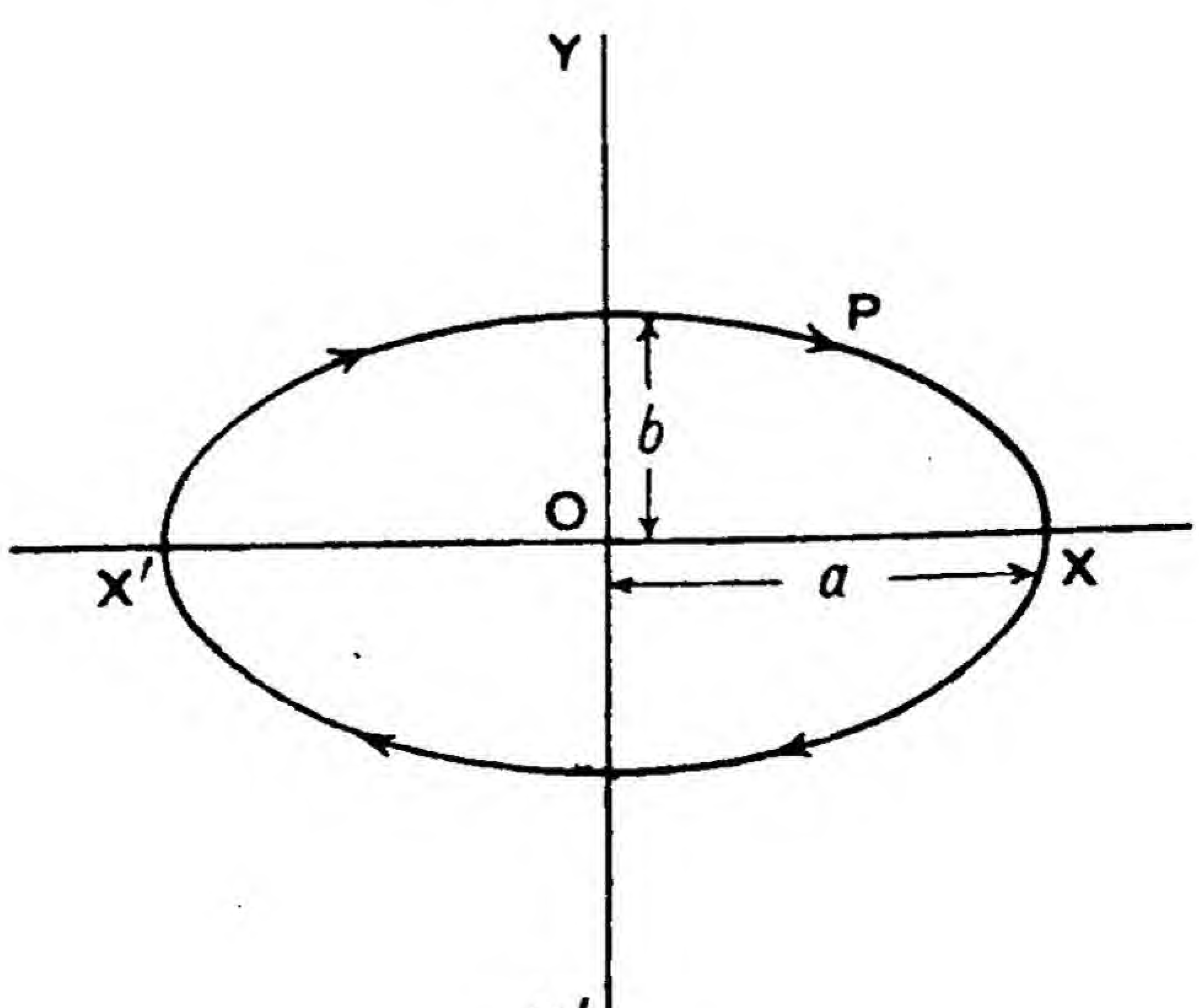


Fig. 192.

Adding these two equations, we have

$$\frac{y^2}{b^2}+\frac{x^2}{a^2}=1$$

which is the equation to an ellipse whose semi-axes coincide with the axes of co-ordinates and are of length a and b (Fig. 192). The particle describes this ellipse once in each period of the component S.H.M.'s. If the amplitudes are equal, the ellipse becomes the circle

$$y^2+x^2=a^2$$

Case 3: The S.H.M.'s are π out of phase and the amplitudes are unequal.

Let them be represented by the equations

$$x=a \sin \omega t$$
$$y=b \sin (\omega t+\pi)$$
$$\text{or } y=-b \sin \omega t$$

Dividing the two equations we have

$$\frac{y}{x}=-\frac{b}{a}$$

which is the equation to a straight line making an angle $-\tan^{-1}\frac{b}{a}$ with

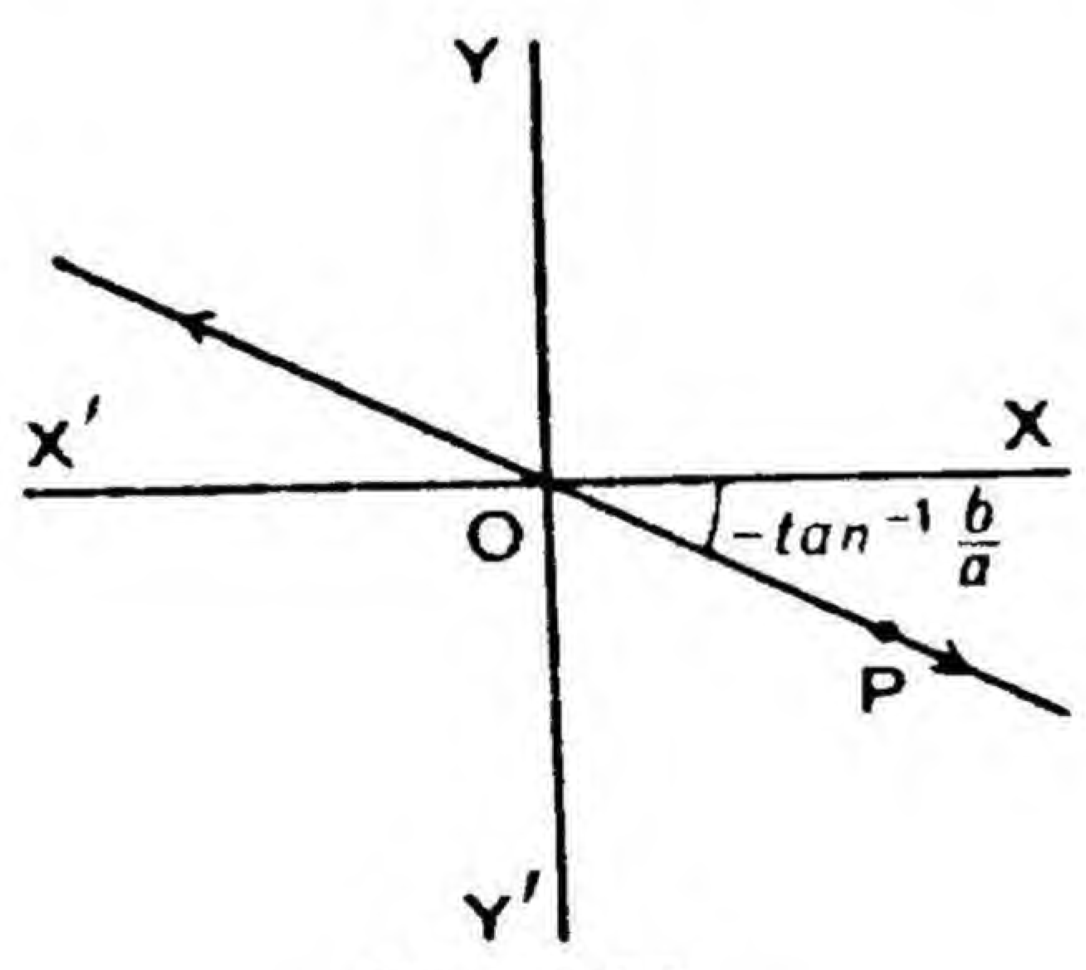


Fig. 193.

the x axis (Fig. 193). The particle describes one to-and-fro motion along this line in the common period of the S.H.M.'s.

We are now armed with the technical equipment we need for further conquests. Let us push on into the unknown, following first the signpost labelled interference.

EXAMPLES ON CHAPTER XII

1. Give an elementary account of the wave theory of light, and prove by its means that the focal length of a concave mirror is equal to half its radius of curvature. *(O. and C.)*

2. What are the principal characteristics of wave motion ?

By considering the refraction of a plane wave at a plane surface, show how the laws of refraction of light can be explained on the wave theory. *(O. and C.)*

3. By considering the case of a plane wave incident obliquely on a plane refracting surface, explain the refraction of light on the wave theory.

An air cell consisting of two vertical parallel plates of glass separated by an air film can be rotated about a vertical axis in a liquid. Explain how you would use the apparatus to find the refractive index of the liquid, and prove any formula you would use. *(Camb. Schol.)*

4. A plane wave front in water is incident on the plane face separating the water from air at an angle of (a) 30°, (b) 60°. Use Huygens' Principle to construct the refracted wave front in each case. Does this lead to an explanation of the impossibility of refraction under certain conditions ?

5. Explain the principle of superposition and secondary wave as used in the wave theory of light and show how Huygens deduced the laws of refraction by means of the theory.

How is total internal reflection explained by Huygens' theory ? *(N.U.J.B.)*

6. Light from a point source falls on the plane surface of a transparent medium of refractive index μ. Discuss the ensuing phenomena from the point of view of the wave theory of light. *(O. and C.)*

7. On the basis of the wave theory deduce the laws of reflection of light in the case where the velocity after reflection differs from the velocity before reflection. *(Camb. Schol.)*

8. Employ the wave theory of light to account for the dispersion of light by a transparent prism.

What information about the optical properties of a substance may be gained from measurements relating its refractive index with the wave-length of light it transmits ? *(Camb. Schol.)*

9. Explain, on the basis of the wave theory of light, how a lens forms an image of a point on its axis.

Use the ideas of the wave theory to deduce an expression for the focal length of the lens, pointing out very clearly the assumptions made. Point out which of these assumptions are not true in practice, and in what way an actual lens may differ from the simple theoretical lens as a result. *(Camb. Schol.)*

10. Apply the principles of the wave theory to find the relation between the focal length of a thin lens, the radii of curvature of the surfaces, and the refractive index of the material.

If a convex lens of focal length 15 cm., made of glass of refractive index 1·52, is totally immersed in a liquid of refractive index 1·35, how will its focal length be affected ? *(Camb. Schol.)*

11. Describe an experiment which you regard as proving decisively that light consists of a wave motion.

Indicate how a convex lens converges a parallel beam of light waves to a focus. *(Oxford Schol.)*

12. Apply the principles of the wave theory to establish the formula

$$\frac{1}{f}=(\mu-1)\left(\frac{1}{r}-\frac{1}{s}\right)$$

for a thin lens.

A ray of light is incident in a direction parallel to the axis on one face of a thin double convex lens of refractive index 1·5, and, after two internal reflections, emerges from the second face. Show that it will cut the axis at a point at a distance $f/7$ from the centre of the lens. *(Camb. Schol.)*

13. Discuss the phenomenon of interference, stating what effects can be observed and what conditions must be satisfied in order to observe them. Illustrate your answer from as many branches of physics as possible.

14. A " seconds " pendulum, whose periodic time is two seconds, has an amplitude of 10·0 cm. Find its maximum velocity, maximum acceleration, its velocity when the displacement is 5·0 cm., and its displacement when its velocity is half the maximum value.

15 Describe precisely the type of motion which is represented by the equation

$$\xi = A \sin 2\pi\left[\frac{t}{\tau} - \frac{x}{\lambda}\right]$$

where ξ is the displacement of the medium at a distance x from the origin, t is the time and A, τ, λ, are constants. Hence show how two wave trains proceeding in opposite directions may in certain circumstances give rise to a set of " stationary waves." Give some examples of this type of wave motion. (*Camb. Schol.*)

16. Discuss the combination of two S.H.M.'s of the same period along two lines at right angles to each other. What type of motion would you expect, if the two periods are not quite equal ? Can you devise a way of illustrating such a motion with a simple pendulum ?

Chapter XIII

INTERFERENCE

120. THE REALISATION OF INTERFERENCE WITH LIGHT

How are we to obtain interference fringes with light ? What sort of apparatus must we set up ? Is it any use taking two pin-holes illuminated by white light or sodium light and looking for fringes on a screen placed a few centimetres away ? We may say at once that an experiment of this kind meets with no success and this is hardly surprising, as it bears only a superficial resemblance to the conditions under which interference was obtained with water waves and sound. Let us examine those conditions a little more closely, so that we may know the precise experimental arrangement most likely to lead to success in light. We have seen that interference can be obtained with two separate tuning forks in sound ; but it is common knowledge that it is not obtained with two separate violins or two separate singers emitting the same note. What is the reason for this ? Let us imagine that we are equidistant from the two singers ; we should expect to be at a place of maximum sound and the amplitude should be double and so the loudness four times that with a single singer. But we only hear double the sound of one singer. If we move off to one side, we never get into a patch of silence. Is the explanation on these lines ? We are only at a place of maximum sound, if the two sources are in phase. But if their phase difference is continually changing, now zero, now π, and so on many times a second, we shall be now at a position of maximum sound, now at place of silence and so on, these changes succeeding one another many times a second. So we shall hear a sound of uniform loudness and the same thing will be true of any other place. To put it in another way, the continual changes in phase difference between the sources causes a continual movement of the interference fringes, so that they cannot be detected. Whether our explanation be true or not, it is a fact that interference cannot be detected with independent sources such as violins or voices in sound. It is possible with tuning forks, because their phase difference is constant, if their frequencies are the same. But it is quite possible to obtain interference with any type of sound source, if two sources are derived from one source. This can be done by causing a whistle to be placed near a mirror, when interference is obtained between the direct waves and those reflected from

the mirror. In this case both waves come from the same source; the two sources derived from the single source are the source itself and its image in the mirror. It is evident that the phase difference between two such sources is always constant, which is why they can produce interference. Such sources are called **coherent**. The two sources producing interference in water waves satisfy this condition. So the first condition essential to the realisation of interference in light is that **the two sources must be coherent; their phase difference must be constant**; this implies that they must be derived from a common source in practice..

We have seen that the fringe width with two point sources decreases as the wave-length decreases. Since white light consists of different colours, that is, of different wave-lengths on the wave theory, we shall have a set of fringes of different width for each colour, if we use white light. This will produce blurring and may make it difficult to detect the interference, so we had better begin by using light of one colour, or monochromatic light, as it is called. Then we shall get bright and dark fringes.

We have also seen that, if light is waves, its wave-length is likely to be very small, otherwise it would show appreciable diffraction. Since the fringe width gets less, the less the wave-length, the fringes may be so close together that we shall not be able to see them. We had better do all we can to make them far apart. So we must put the sources close together and put the screen as far as possible from them, since the distance apart of the fringes gets bigger as the distance apart of the sources decreases and as their distance from the screen increases (Art. 115). Finally we could replace the screen and the naked eye by an eyepiece, if necessary. To sum up, then, we **can only hope to get interference fringes in light if we use coherent sources emitting monochromatic light, put the sources as close together as possible, and put the screen of observation as far from them as possible.**

We are now in a position to design our interference apparatus. But it is not necessary, because it has already been set up for us! In fact, the phenomenon had already been discovered a hundred years before it was recognised! This frequently happens in the course of scientific investigation. Nature whispers, but it may be that no one is sufficiently curious or sensitive to truth to perceive the whisper. Or it may be that the message is too mysterious to be understood. One of the things which helped Röntgen to discover X-rays was the fogging of photographic plates wrapped in black paper near a discharge tube. Röntgen followed up this whisper of Nature and other clues and discovered X-rays. After he had announced his discovery, an English physicist said that he had noticed the fogging of photographic plates, but he had just moved his plates further away. He was not sensitive to the whisper of Nature, so he saved his plates but he lost the X-rays! In the present case the whisper of Nature had been noticed by Grimaldi in that he had discovered

the equally spaced dark and bright bands inside the shadow of a narrow obstacle cast by an illuminated slit. This showed that all was not well with rectilinear propagation, but Grimaldi and his successors were not able to fit this clue into any rational scheme. But a hundred years later, Young saw that these equally spaced dark and bright bands were interference fringes, the two sources being the light which bends round each edge of the narrow obstacle. We shall see shortly that mathematical theory leads to the expectation of equally spaced bands under the conditions of this experiment, where we have two sources very close together and the screen a long way from the sources. We see that the experiment accidentally reproduces all the conditions we have discussed, as the two sources are coherent being derived from the single illuminated slit. Young clinched the argument by preventing light from one edge of the obstacle from reaching the screen and the fringes disappeared and were replaced by a faint uniform illumination. Thus interference has been obtained with light and reveals the characteristic that **light added to light produces darkness**, the dark fringes disappearing when only one beam of light is used. A further confirmation that the bands were interference fringes was made by showing that no fringes were obtained when a wide obstacle was used. This is because the light does not bend round into the geometrical shadow sufficiently to enable the two beams to cross and interference only occurs where two wave trains cross. We must not be too hasty in accepting this evidence however. The interfering beams are rather peculiar ; they are not light rays in the ordinary sense, they are rays which have bent into the geometrical shadow. But the enthusiastic supporter of the wave theory will reply that this is diffraction, another property we seek in light, so why worry ? But it would be just as well to get interference with ordinary rays of light, and we must do another vitally important thing ; we must measure the wave-length of light in numbers and see if different colours do have different wave-lengths as we anticipate. We shall now proceed to various ways of doing this and putting the phenomenon of interference in light beyond all dispute, noticing on the way that Young also saw that Newton's rings could be regarded as a case of interference, the two interfering wave trains being produced by reflection at the lower surface of the lens and the upper surface of the glass plate. We shall defer a consideration of this and similar cases to a later article.

121. YOUNG'S SLITS

The first deliberate attempt to produce interference in light was due to Young and is known as Young's Slits and a horizontal cross-section of his apparatus is shown in Fig. 194. A single vertical slit S is illuminated by a sodium lamp L or some other source of monochromatic light. Two other parallel slits S_1 and S_2 are placed in front of and parallel to S.

PLATE V

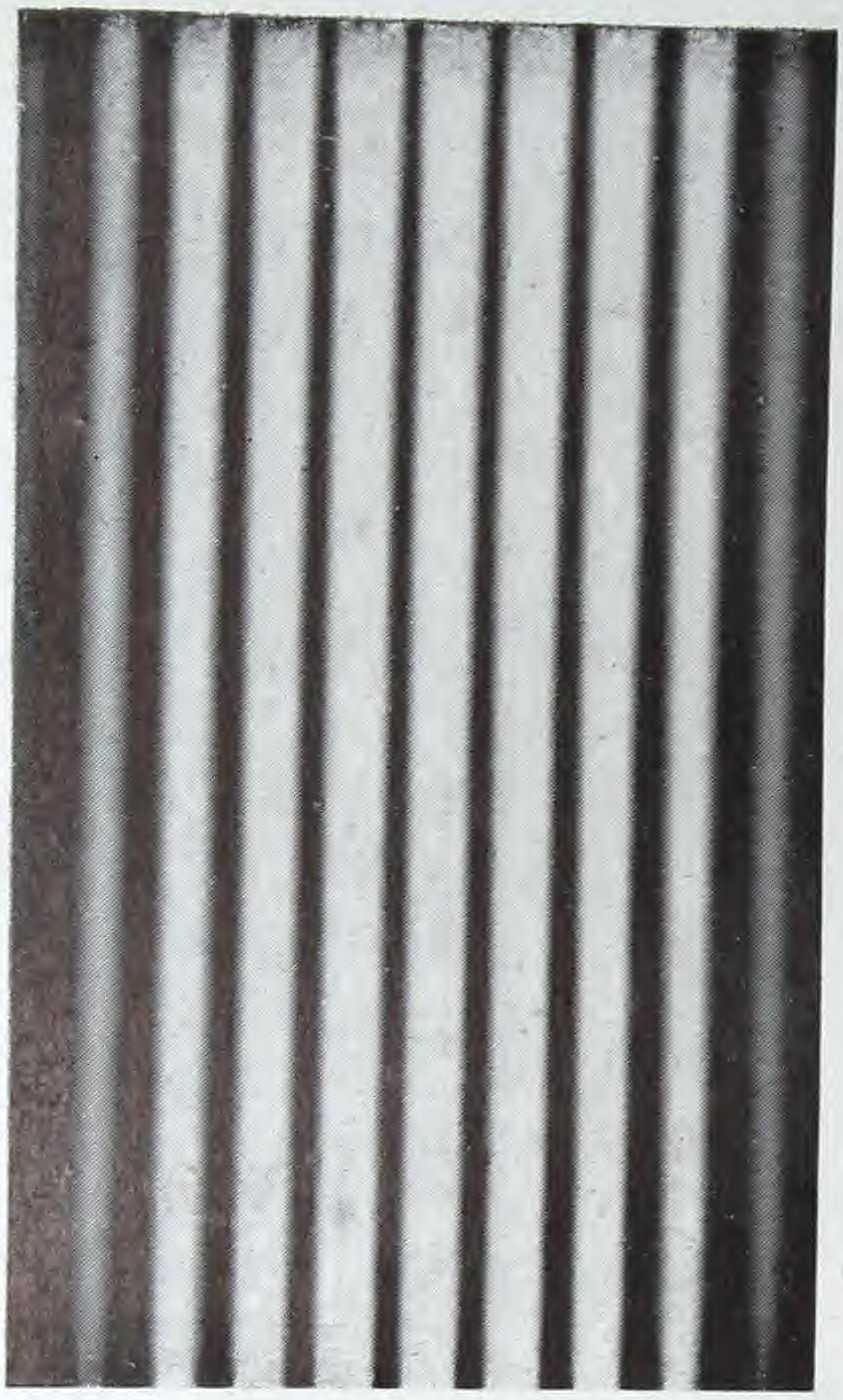


Fig. 1. Interference fringes produced by Young's Slits, using red light.

(*J. W. Cottingham*)

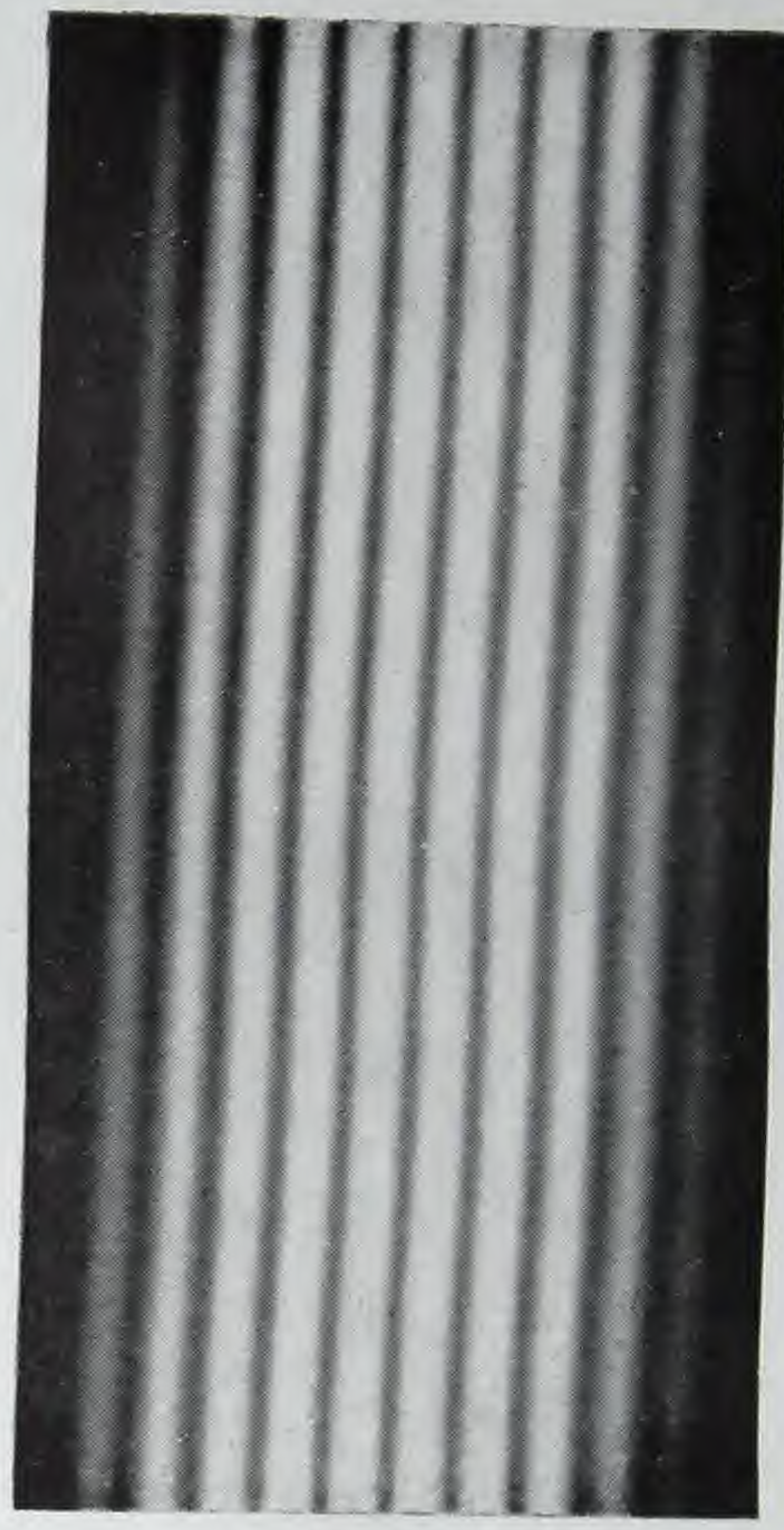


Fig. 2. Interference fringes produced by Young's Slits, using blue light, showing the smaller fringe width due to the smaller wave-length of blue-light.

(*J. W. Cottingham*)

Fig, 3. Plane waves exhibiting a small amount of diffraction in passing through an aperture large compared to their wave-length.

(*J. W. Cottingham*)

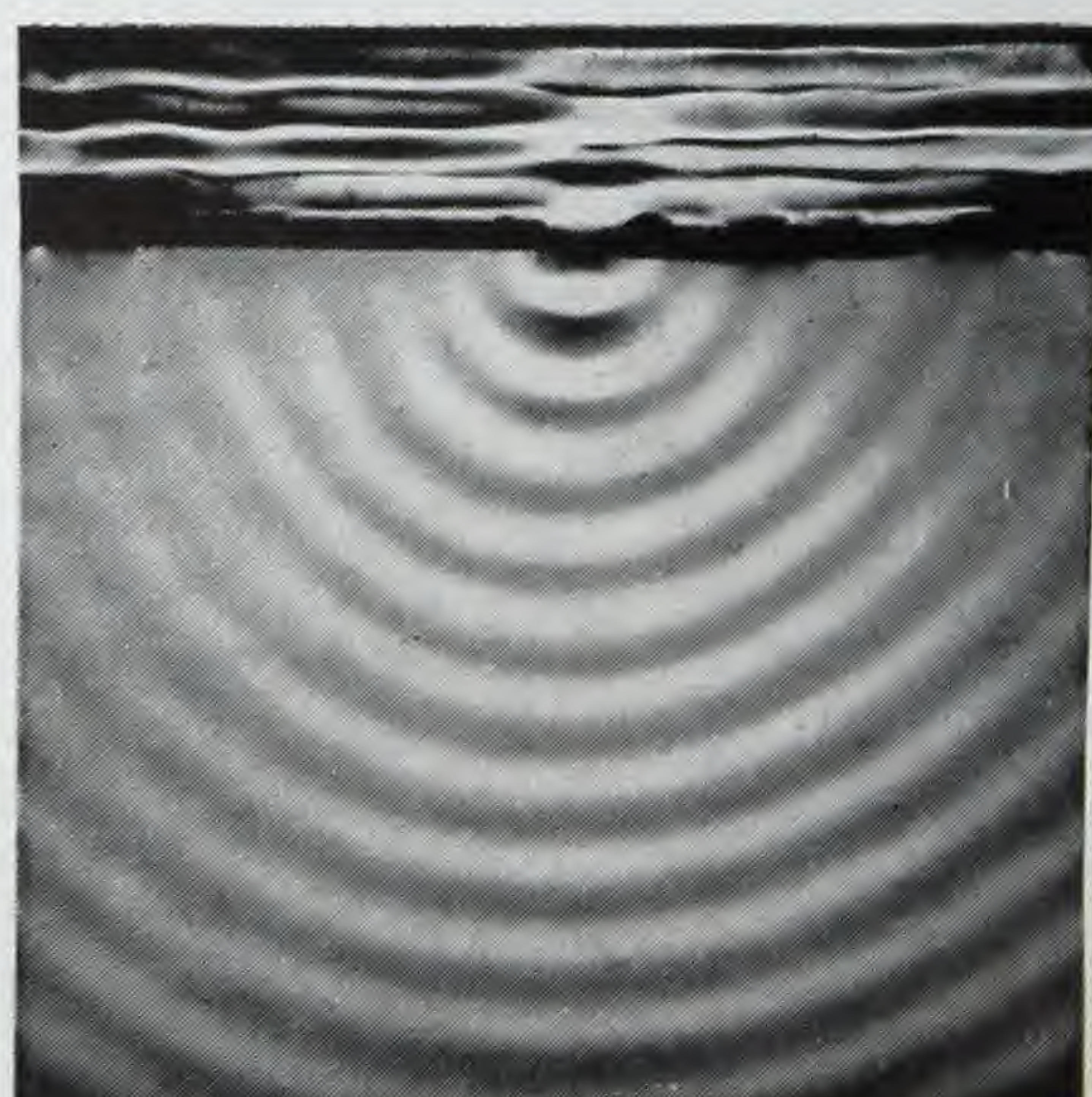


Fig. 4. Plane waves exhibiting a large amount of diffraction in passing through an aperture of about the same size as their wave-length.

(*J. W. Cottingham*)

They are accurately parallel to each other and are about 0·3 mm. apart, each slit being 0·03 mm. wide. Light from S falls on to the two slits and passes through them spreading out as shown and interference should be visible in the region in which the light from S_1 and S_2 crosses. We

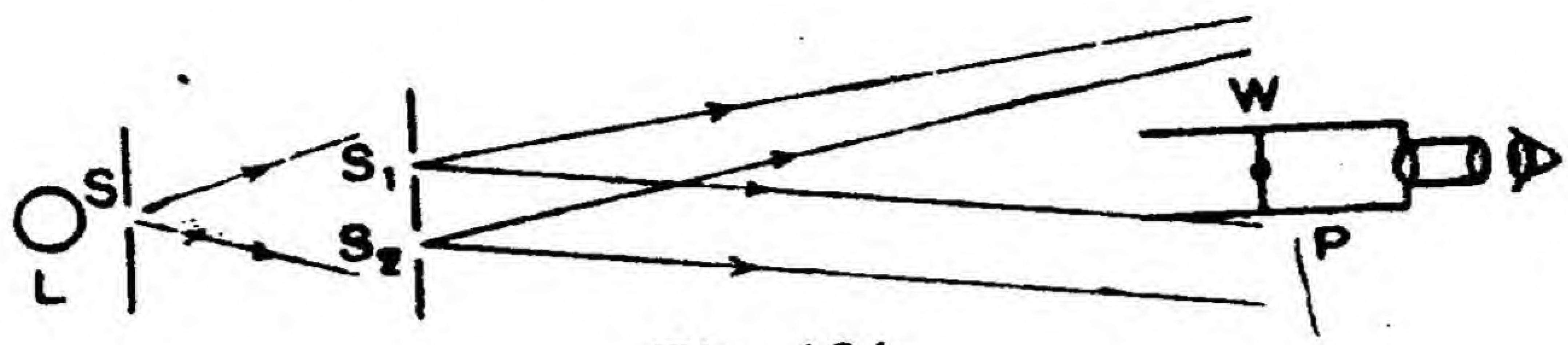


Fig. 194.

have already seen (Art. 115) that two point sources of waves produce fringes which are hyperboloids of revolution about the line joining the sources as axis and the intersection of these hyperboloids with a vertical screen will be a set of hyperbolæ. So we should expect to see a set of alternate dark and bright hyperbolæ on such a screen or in the travelling eyepiece P. As the sources are so close together and the screen is so far from them, about 30 cm. in practice, we see a set of equally spaced alternate dark and bright vertical fringes. The reader will notice how this experiment satisfies the conditions mentioned in the previous article. The two interfering sources S_1 and S_2 are derived from the single source S and so their phase difference is constant. To be more precise, the wave emitted by a single atom of the source L spreads out through S, and covers S_1 and S_2, and a portion of this single wave front spreads out from S_1 and S_2, and these portions are always in a condition to interfere. Monochromatic light is used, the sources are close together, and the screen is a good distance from them. To satisfy ourselves still further that the fringes seen are due to interference, we must see if they exhibit the variation with conditions shown by the fringes obtained with water waves (Art. 115). If the screen or travelling eyepiece is moved further away, the fringe width increases and, if the experiment is repeated with two slits closer together, the fringe width at a given distance from the sources also increases (Plate V). Finally, the fringes do disappear if one of the sources be covered up showing that it is light *added* to light which produces darkness. Thus it is certain that the fringes are genuine interference fringes. The fringes are closer together for blue than for red light, showing that the wave-length of blue light is less than that of red light. We must now consider the theory of the fringes, first to verify that equally spaced fringes are what we should expect and secondly to see if we can get the wave-length of light from this experiment.

Let S_1 and S_2 (Fig. 195) represent two point sources a distance s apart emitting waves of wave-length λ in the same phase and let interference be produced on a screen at a distance d from the plane of the two sources. The central bright fringe is produced at O on the right bisector of the two sources and the first bright fringe occurs at the point A where $S_2A - S_1A$

$=\lambda$, the second at B where $S_2B-S_1B=2\lambda$. If P is at the n^{th} bright fringe,

$$S_2P-S_1P=n\lambda$$

Now
$$S_2P^2-S_1P^2=\left\{d^2+\left(x_1+\frac{s}{2}\right)^2\right\}-\left\{d^2+\left(x_1-\frac{s}{2}\right)^2\right\}$$
$$=2x_1s$$

$$\therefore\ (S_2P-S_1P)=\frac{2x_1s}{S_2P+S_1P}$$

In all the cases of interference we shall consider, $s=0{\cdot}3$ mm., $d=30$ cm., and $x_1=5$ mm. are typical values, and so the error in putting $S_2P+S_1P=2d$

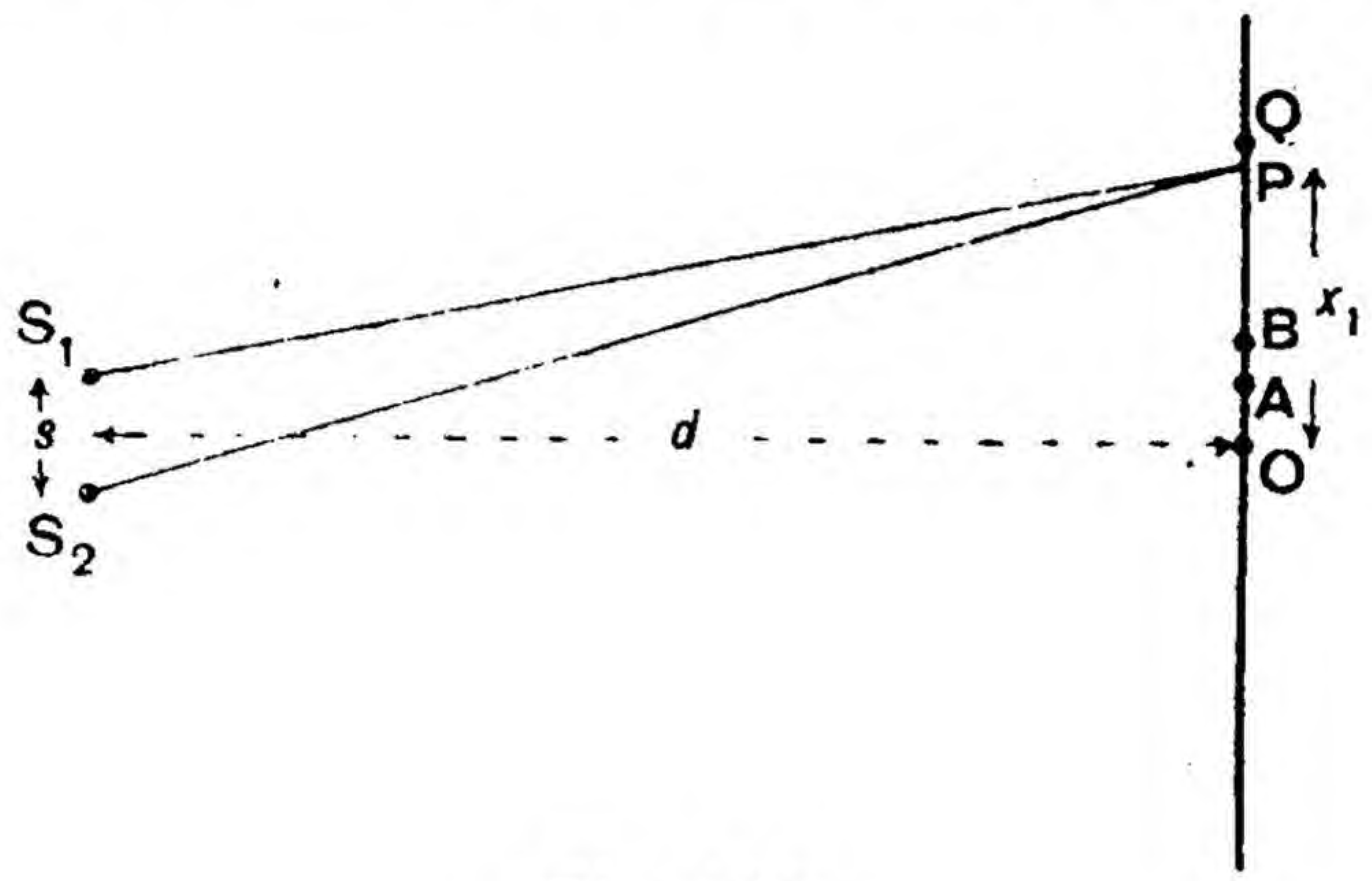


Fig. 195.

is less than 1 per cent., which is well within the accuracy with which the fringe-widths can be measured. Making this approximation, we have

$$S_2P-S_1P=\frac{x_1s}{d}$$

Since P is at the nth fringe,

$$S_2P-S_1P=n\lambda$$

$$\therefore\ \frac{x_1s}{d}=n\lambda$$

$$\therefore\ x_1=\frac{nd\lambda}{s}$$

If the $(n+1)^{th}$ bright fringe is at Q, a distance x_2 from O,

$$x_2=\frac{(n+1)d\lambda}{s}$$

and the fringe-width ω is given by

$$\omega=x_2-x_1$$

$$\therefore\ \omega=\frac{d\lambda}{s}\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (71)$$

This is independent of the distance of the fringes from O and so our

observation of equally spaced alternate dark and bright fringes is confirmed theoretically.

This result also shows us how to measure the wave-length of light from these fringes. We measure the fringe-width by means of the travelling eyepiece, finding the distance between as large a number of dark fringes as we can see clearly and dividing this distance by the number of fringes. We work with the dark rather than the bright fringes because they are the narrower, so that it is possible to set the vertical cross-wire of the eyepiece more accurately on the centre of a dark than a bright fringe. The distance d from the plane of the slits to the cross-wires, W, of the eyepiece is measured by a metre rule and finally the parallel slits are placed under a travelling microscope and their distance apart s is measured. The wave-length is calculated from equation (71). A typical set of results for sodium yellow light is $s=0{\cdot}03$ cm., $d=30{\cdot}0$ cm., $\omega=0{\cdot}59$ mm., whence $\lambda=5{\cdot}90\times10^{-5}$ cm.. The wave-length of red light turns out to be about 8×10^{-5} cm. and of blue light $4{\cdot}5\times10^{-5}$ cm. This very small value of the wave-length of light is an additional reason for beginning to believe in the wave theory, for it is just what is needed to explain away rectilinear propagation.

The reader may object that the theory has always dealt with point sources, while the experiment uses two slits. But the two slits may be regarded as a set of pairs of point sources, any one pair being in the same horizontal line. Each pair will produce equally spaced vertical dark and bright fringes and the different sets will fit over one another and make the bright fringes brighter. The sole effect of substituting slits for points is to get brighter fringes. But another objection to accepting the fringes as being due to interference was made at the time of Young's experiment. This may seem surprising in view of the way in which both the qualitative and quantitative facts about the fringes fit the theoretical predictions. But the objection was based on the ground that, if light travels in straight lines, there can be no crossing of the light which emerges from S and then goes through S_1 and S_2. The interference fringes have not been obtained with ordinary rays of light, but only with light which has bent round the corner of a narrow aperture. There is certainly something in this objection and Fresnel devised two experiments which succeeded in disposing of it by arranging for the crossing of rays of light which had travelled in straight lines. We shall discuss these and a number of other cases of interference with two point sources and the reader will see that, in every case, the fundamental feature of the apparatus is a method for making two light sources out of a single source, so that their phase difference shall be constant.

122. FRESNEL'S BI-PRISM

Fresnel's first arrangement for producing two sources out of one is the bi-prism, which is a single prism, one of whose angles is of the order of

179° 20′ and the other two are 20′. It is used as two prisms each of refracting angle 20′. A horizontal cross-section of the apparatus is shown in Fig. 196, in which S represents a vertical slit, illuminated with monochromatic light, placed some 5 cm. from the bi-prism B, whose two refracting edges are vertical. Each half of the bi-prism produces an image of the slit S and these two images S_1 and S_2 can be seen with the naked eye, if it is placed on the other side of the bi-prism from the slit S. The distance of S from the bi-prism is adjusted until the images are very close together (about 0·3 mm. is a suitable value) and the slit S is rotated

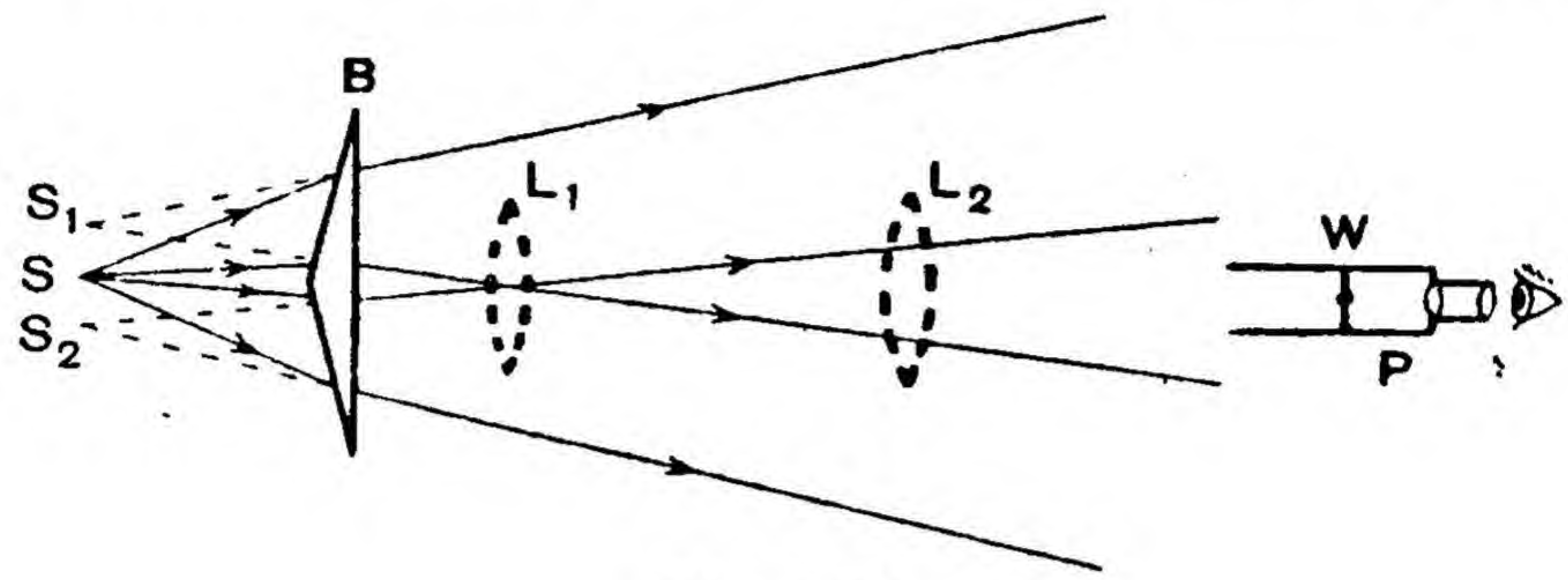


Fig. 196.

in its own plane until the two images are parallel to one another. Equally spaced interference fringes can then be seen on a vertical screen placed about 30 cm. from the bi-prism or they can be observed in a travelling eyepiece P. The fringe-width increases if the eyepiece is moved further from the bi-prism or if the sources S_1 and S_2 are moved closer together by moving the slit S nearer to the bi-prism and they can be made to disappear by covering up one half of the bi-prism. There can be no doubt that these are interference fringes, for they satisfy all the qualitative and quantitative requirements of the theory and the beams of light are made to cross by refraction instead of by a process, diffraction, which was not understood in Young's time.

The wave-length of light can be measured in the following way. The fringe-width, ω, is found as in Young's slits by the travelling eyepiece. The distance apart s of the sources and their distance d from the cross-wires of the eyepiece are found by putting a lens in the position L_1 between S and W and adjusting it so that images of S_1 and S_2 fall on the cross-wires of the eyepiece and so can be seen in focus. The distance apart s_1 of these images is measured by the eyepiece and also the distance a from the lens to the cross-wires. The lens is now moved to the second position L_2, which will bring sharply focussed images of S_1 and S_2 in the eyepiece. The distance s_2 between these images is measured as before and the distance b from the lens to the cross-wires. The reader should have no difficulty in proving that $s=\sqrt{s_1 s_2}$ and $d=a+b$. The wave-length of light can then be calculated by substituting these values of ω, s, and d in equation (71). The value obtained for sodium yellow light, for example, agrees within the limits of experimental error with that obtained from Young's slits.

123. FRESNEL'S INCLINED MIRRORS

Fresnel's second arrangement for producing interference is shown in horizontal cross-section in Fig. 197. A vertical slit S illuminated with monochromatic light is placed just in front of two mirrors M_1 and M_2, which are made of black glass, so that only reflection at the front surface

Fig. 197.

takes place. The line of intersection of the front surfaces of the two mirrors is adjusted to be vertical and the mirror M_1 can be rotated about this line as axis. An eye is placed so that the images S_1 and S_2 of the slit S in the mirrors can be seen; when M_1 and M_2 are in the same plane, S_1 and S_2 fuse into a single image. M_1 is then rotated through the angle needed to produce two images S_1 and S_2 about 0·03 cm. apart. The angle between the mirrors is then about 10′. Interference fringes can then be seen in the travelling eyepiece P which is placed in the region where the light from S_1 and S_2 crosses, or the fringes can be seen on a screen placed in the same region. These fringes are much harder to obtain than those with the bi-prism, since only a small fraction of the light emerging from the slit S is reflected from the mirrors and the bright fringes are of small intensity. Great care must be taken to exclude all stray light, including light which comes direct from the source S itself, and the experiment must be carried out in a dark room. When the fringes have been obtained and their contrast has been adjusted to be a maximum, the wave-length of light is measured as with the bi-prism and the results obtained for sodium yellow light, for example, agree within the limits of experimental error with those found from the previous two methods.

124. BILLET'S SPLIT LENS

So great was the interest aroused in the phenomenon of interference, that many other attempts were made to produce interference with two point sources. These varieties are simply different ways of making two sources out of one. Billet's split lens is one of these and two images of the illuminated slit S (Fig. 198) are produced by splitting a lens into two parts along a diameter of its circular outline and moving them a

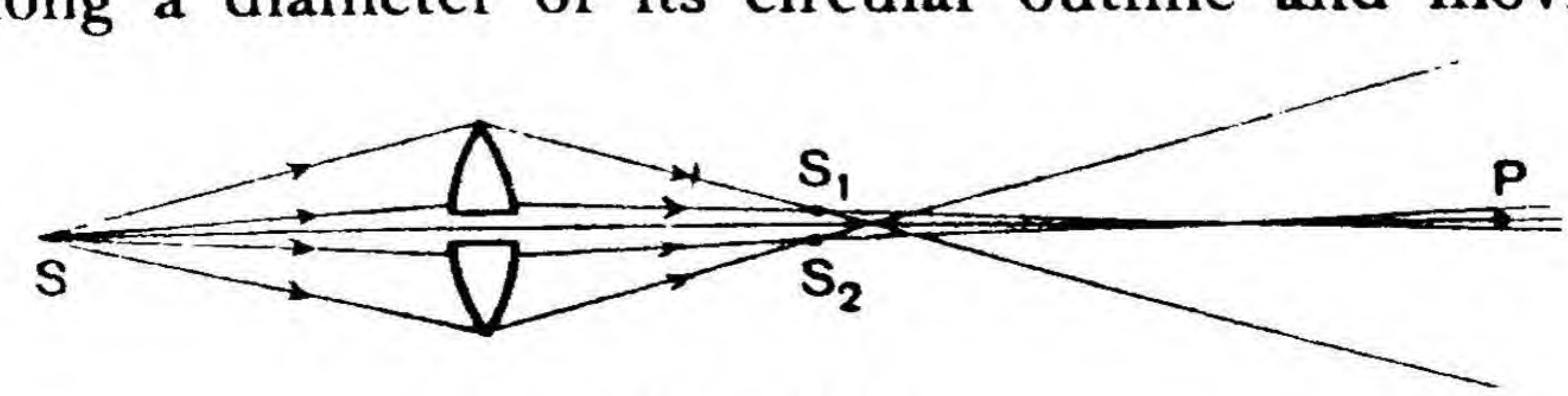


Fig. 198.

small distance apart, until the images are a suitable distance from one another as seen by the naked eye. Fringes can then be seen on a screen or in an eyepiece placed at P in the region where the light from the two sources crosses. An alternative way of using the split lens is shown in Fig. 199 and produces one source behind the other. Since the two point sources produce fringes which are hyperboloids of revolutions about S_1S_2 as axis, the intersection of these hyperboloids with a screen normal to the axis of the lens is a set of circles and circular fringes are produced in this case.

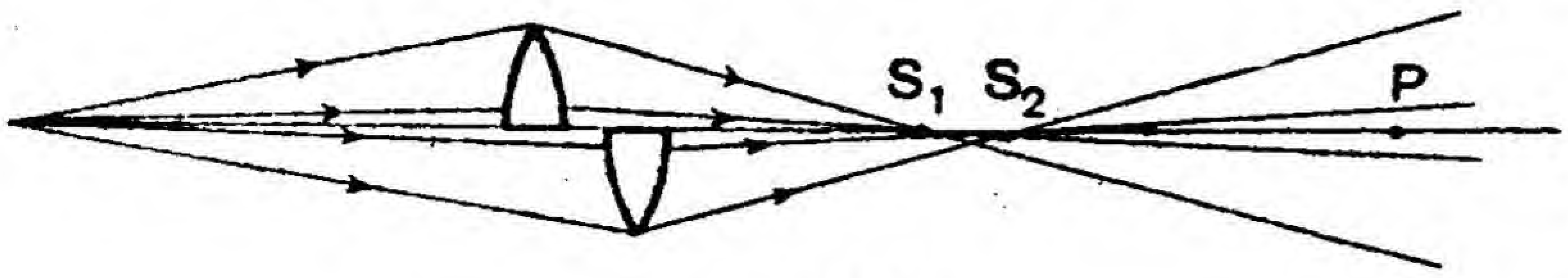


Fig. 199.

125. LLOYD'S SINGLE MIRROR

A vertical slit S (Fig. 200) illuminated with monochromatic light is placed close to the front surface of a glass plate M, which is adjusted so that its front surface is in a vertical plane. An image S_1 of S is produced

Fig. 200.

by reflection in the front surface of M and interference occurs between the two sources S and S_1, in other words, between the direct and reflected rays from S. The source S is moved towards M until the slit and its image S_1 are about 0·03 cm. apart as seen by the naked eye, and the slit is then rotated until it is parallel to its image. Interference fringes can then be seen either on a vertical screen or in a travelling eyepiece placed in the region of P, where the direct and reflected rays cross. The wave-length of light can be measured in the same way as with the bi-prism and the values obtained agree within the limits of experimental error with those obtained from the previous methods. These fringes are quite easy to obtain; the amplitudes of the direct and reflected waves are about equal, since the light falls on the front surface of the glass plate so obliquely that a large percentage of the incident light is reflected. This equality of the amplitudes of the two wave trains is important in practice, otherwise the dark fringes are nearly as bright as those of maximum intensity and the contrast between the dark and bright fringes is so poor, that it is difficult to see them at all. There is no need to worry about light reflected from the back surface of the plate, as the image of the slit in the back surface is too far away from either S or S_1 to produce fringes wide enough to be seen.

The fringes produced by Lloyd's single mirror are different from those obtained by the previous methods, in that only one half of the field can be seen. To be quite accurate rather less than one half of the field is

visible, the central fringe of zero path difference being invisible, because it lies on the right bisector of S and S_1 which is on the surface of the plate M produced and no reflected light can reach this point. But the fringe of zero path difference can be made visible by moving it away from the right bisector of S and S_1 by the insertion of a thin piece of mica in the path of the direct beam. If the mica has a thickness t, its insertion replaces t cm. of air by t cm. of mica, which is equivalent to nt cm. of air, where n is the refractive index of mica. So the path of the direct beam is increased by $(n-1)t$ due to the insertion of the mica so the place of zero path difference is moved upwards by $\frac{(n-1)t}{\lambda}$ fringes.

If a piece of mica of suitable thickness is chosen, the fringe of zero path difference can be made visible. When this is done, **it is found to be a dark fringe** and not a bright one, as would be anticipated. Incidentally there is no difficulty in identifying the central fringe ; it is only necessary to replace the monochromatic light by white light, when the central fringe becomes the only white fringe visible, because it is the only place which is a maximum for all wave-lengths. But in this case, that fringe was black instead of white ! There can only be one explanation of this, namely that **the reflected ray suffers a π change of phase on reflection**. That is, a crest is reflected as a trough and vice-versa. If this is the case, then a crest from S will arrive at the same time as a trough from S_1 at the point of zero path difference and a black fringe will be produced. In this case the fringes are bright or dark according as the path difference is $(2n+1)\frac{\lambda}{2}$ or $n\lambda$, where n is an integer.

This important result can be verified and extended theoretically by using the principle of the reversibility of rays of light. If a ray of amplitude a is incident on a plane surface of a more dense medium, a reflected ray of amplitude ar and a refracted ray of amplitude at are produced, where r and t are the reflecting and transmitting coefficients of the medium respectively (Fig. 201). If the reflected and refracted rays are reversed, they should produce the incident ray only. The reflected ray will produce a transmitted ray in the more dense medium of amplitude art, while the refracted ray will produce a reflected ray in the more dense medium of amplitude atr', where r' is the reflecting coefficient for light reflected in the more dense medium. But there must be no ray in the more dense medium and so

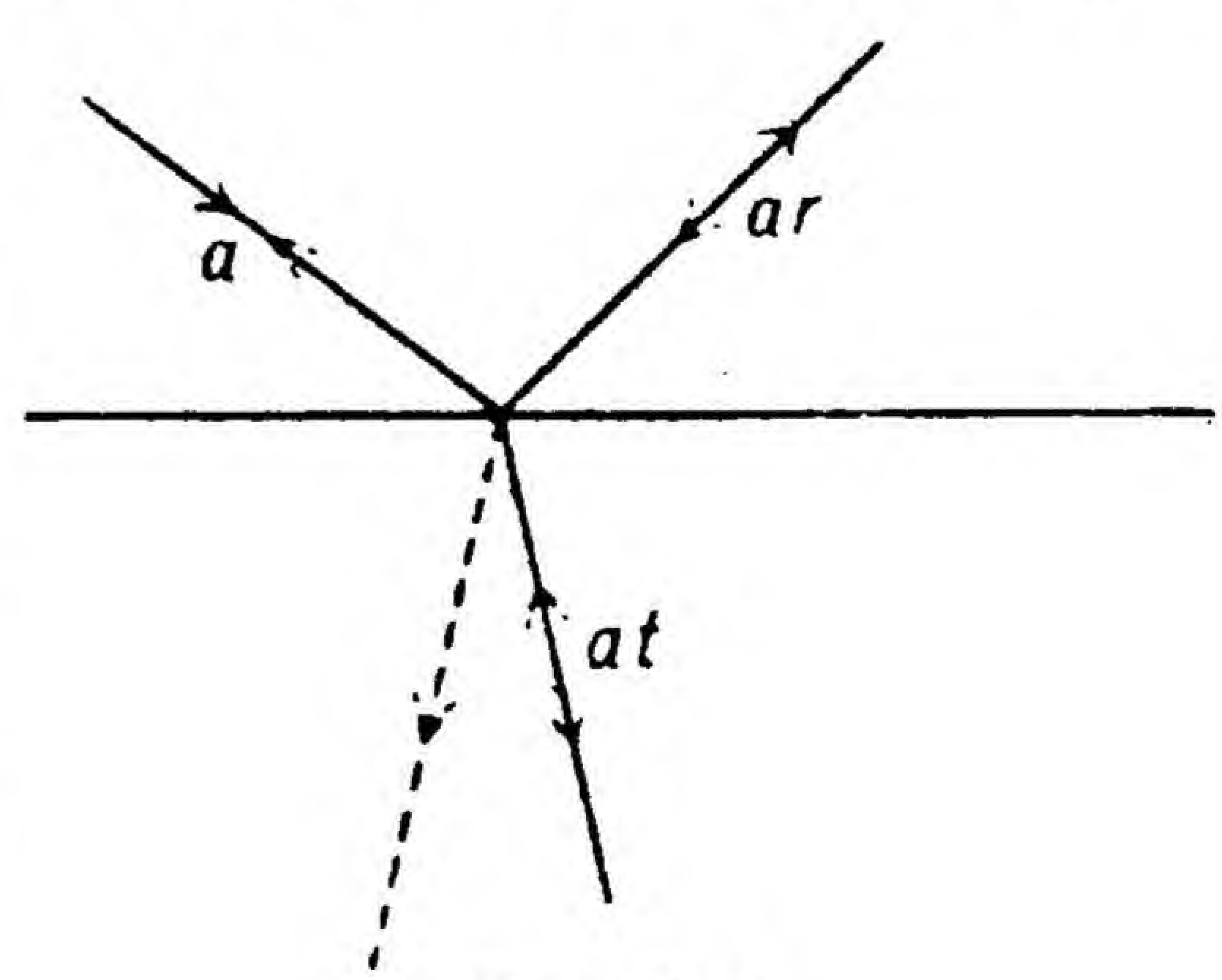


Fig. 201.

$$art+atr'=o$$
$$\therefore r=-r'$$

But we know from Lloyd's single mirror that there is a π change in phase when light in air is reflected at the boundary of a more dense medium and so this result proves that there is no phase change when light in a more dense medium is reflected at the boundary of air.

We shall conclude this account of the ways of producing interference fringes with two point or slit sources by mentioning some features, common to all the cases, which must be observed if good fringes are to be obtained in practice. Firstly the illuminated slit must be strictly parallel to the parallel slits in Young's Slits, or to the line joining the mirrors in Fresnel's mirrors, or to the obtuse edge of the prism in the bi-prism. The adjustment is made in practice by getting it roughly right by eye, when fringes are usually seen. Then the single slit is rotated in its own plane, until fringes of greatest contrast are seen. The necessity for this adjustment is as follows. Considering Young's Slits, suppose one point of the single slit is on the right bisector of the two points in the parallel slits, which are in the same horizontal plane as the one point. Then the fringe of zero path difference is on the right bisector of those two points, since the light from the point in the single slit arrives at those two points in phase. If the single slit is parallel to the double slits, this condition is fulfilled for all the sets of three points into which the slits can be analysed and so the separate fringe systems produced by the sets of points all fit on to one another. But, if the single slit is inclined to the parallel slits, we can easily find a point which is so far off the right bisector of the two points in the parallel slits in the same horizontal plane as itself that the light from it reaches those two points π out of phase. Hence the fringe on the right bisector of these two points is a dark fringe and its set of fringes does not fit on to those produced by the first set of points considered. This type of effect is obviously going to cause blurring of the fringes due to an inexact fit and will spoil the contrast of the fringes. The second adjustment which must be made is to make the slit as narrow as is consistent with having the bright fringes easily visible. If the slit is too wide, the contrast of the fringes decreases. This is due to the fact that a slit of finite width can be analysed into a set of geometrical lines each in different positions, each sending light to the parallel slits and each having a different phase difference at the parallel slits. So each line slit will produce a set of fringes of its own with the central bright fringe in a slightly different position and this inexact fit produces a poor contrast. Finally, there is one other interesting point about the interference fringes which we have considered. We have already seen that interference effects can only be observed at the places where two trains of waves actually cross. But there is a wide region of space in which such crossing occurs, so that it is not necessary to put the observing screen or to focus the eyepiece on any special place. For that reason this class of fringes is called **non-localised fringes**.

126. INTERFERENCE WITH WHITE LIGHT

What will happen if we replace monochromatic light with white light in the above cases of interference? What effects shall we expect to observe? We have already seen that the fringe of zero path difference, often called the central fringe, is the only one which is white and this gives an easy way of finding the central fringe. To see what will happen as we go further out from the centre to places of greater path difference it is best to consider some numerical examples. Taking the limits of the visible spectrum as 8×10^{-5} cm. and 4×10^{-5} cm., we see that the violet will be absent at the place where the path difference is 2×10^{-5} cm., but there is not complete destructive interference for any other colour. So we shall see white light deprived of violet, which is a reddish tint. Further out, where the path difference is 4×10^{-5} cm., the violet will be a maximum while the red is absent, so that we shall see violet here. The central white fringe is bordered by a red fringe, followed by a violet fringe on either side. At a place where the path difference is 16×10^{-5} cm., wave-lengths given by $(2n+1)\frac{\lambda}{2}=16\times 10^{-5}$ will be absent and the only values lying within the visible spectrum are given by $n=2$ and 3 and are $6{\cdot}4\times 10^{-5}$ cm. and $4{\cdot}6\times 10^{-5}$ cm. respectively, lying in the orange and blue. The resultant colour here is white minus orange and blue; red and green will be particularly strong with some violet and the colour will be a greenish yellow. At the place where the path difference is 40×10^{-5} cm., wave-lengths given by $(2n+1)=40\times 10^{-5}$ will be absent and the values of n equal to 5, 6, 7, 8, 9 give wave-lengths lying in the visible spectrum. They are $7{\cdot}3\times 10^{-5}$, $6{\cdot}2\times 10^{-5}$, $5{\cdot}3\times 10^{-5}$, $4{\cdot}7\times 10^{-5}$, $4{\cdot}2\times 10^{-5}$ cm. respectively. When five wave-lengths fairly evenly distributed throughout the spectrum are absent from the light the unaided eye cannot distinguish it from white light, and the greater the path difference the more such wave-length sare absent and the more nearly this becomes true. So we see that with white light we get a white fringe at the point of zero path difference and a few coloured fringes on either side, fading off into uniform white light after some ten fringes. Of course, a spectrometer could detect the difference between white light and the light at the point of path difference 40×10^{-5} cm., for, if the slit of the spectrometer were placed at this point parallel to the fringes, the resulting spectrum would be crossed by five dark bands occupying the positions corresponding to the wave-lengths given above.

127. ACHROMATIC FRINGES

Is it possible to obtain pure black and white fringes using white light? This could be done if it were possible to fix up some arrangement which

would produce fringes whose width was independent of the wave-length. The expression for the fringe-width obtained with two point sources

$$\omega = \frac{d\lambda}{s}$$

shows that this is possible if $\frac{d\lambda}{s}$ can be kept constant. This can be done by making either $\frac{\lambda}{s}$ constant or $d\lambda$ constant or by a combination of the two. It turns out that the first of the three possibilities can be realised with Lloyd's single mirror by using a slit illuminated by a narrow spectrum as the source (Fig. 202). The spectrum is produced by an achromatic lens and the violet end is arranged to be the nearer to the front surface of the glass plate M. Then interference occurs between the source

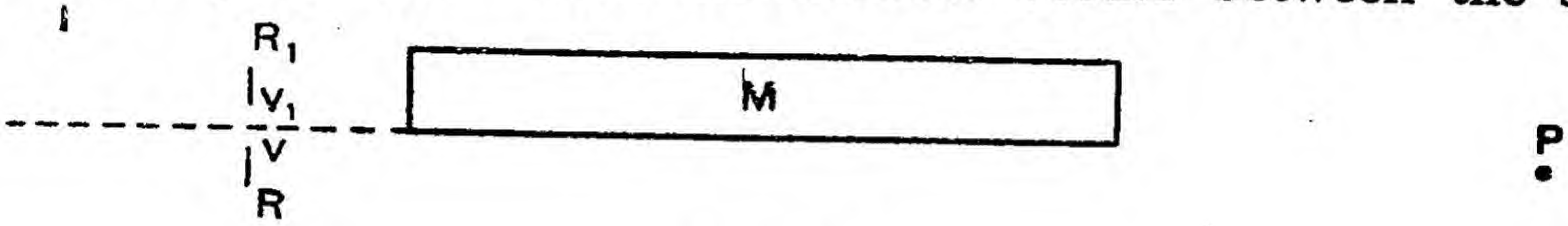


Fig. 202.

VR and its reflection V_1R_1 in the front surface of the glass plate, violet fringes being produced by VV_1, red fringes by RR_1, and so on. If the distance of VR from the surface of the glass plate is correctly adjusted, $\frac{\lambda}{s}$ can be made constant. Fringes are seen in the region of P either on a screen or in an eyepiece and, in practice, the distance of VR from the glass plate is adjusted until the fringes are as nearly black and white as possible. As the distance apart of the sources in monochromatic light is of the order of 0·03 cm., if VV_1 is made equal to this, then RR_1 will be 0·06 cm. at the correct adjustment, which means that the spectrum will be only 0·015 cm. long. Such a narrow spectrum can only be obtained either with a prism of small angle or with a diffraction grating (Art. 142) with about 200 lines per centimetre.

128. THE COLOURS OF THIN FILMS

The success of these experiments deliberately designed to produce interference in light led to increased interest in this topic and redirected attention to the colours of thin films and similar phenomena, which had been observed and discussed by Newton and Hooke. Everyone has seen the exquisite colours produced by a film of oil on the surface of a puddle on the road and many readers will have seen similar colours produced artificially by light reflected from a soap film. Hooke observed such colours in thin flakes of mica and similar natural thin plates and Newton described with typical care and accuracy the rings obtained when a convex lens rests on a plate of glass (Art. 97). We have already seen that Newton's

explanation involved giving light an undulatory nature and Hooke's explanation of the colours of thin plates also involved a wave theory. But no further progress was made for the next hundred years, until Young saw that the essential conditions for interference were coherent sources and, in the case of white light, small path differences. These conditions, together with Newton's observation that the colours in Newton's Rings were replaced by alternate dark and bright rings in monochromatic light, suggested to Young that interference might be the explanation of the phenomena, which all involved thin films and so the possibility of small path differences. Young felt that the key to all these phenomena was that *interference occurs between light reflected from the top and bottom surfaces of the film*; we shall now study this in detail, dealing with some very simple cases which illustrate the principles concerned, and then lead on to the most commonly occurring case with thin films, concluding with Newton's Rings.

If the eye E (Fig. 203) focusses on the point A on the top surface of a thin film, two rays from the point source S can pass through A, the ray 1 which is reflected from the top surface and the ray 2, which is refracted at the top surface and reflected at the bottom surface, to be refracted out of the top surface at A. These rays are coherent, since they have started from the same point of the source, and since they cross at the point A, interference effects can be seen there. Since A′ is the image of A formed on the retina of the eye, the difference in path of the two rays between A and A′ is zero; therefore the path difference of the two rays 1 and 2 at A′ is the same as that at A. If there is a minimum at A, there will be a minimum at A′; we can therefore predict what the eye will record by finding the effect at A.

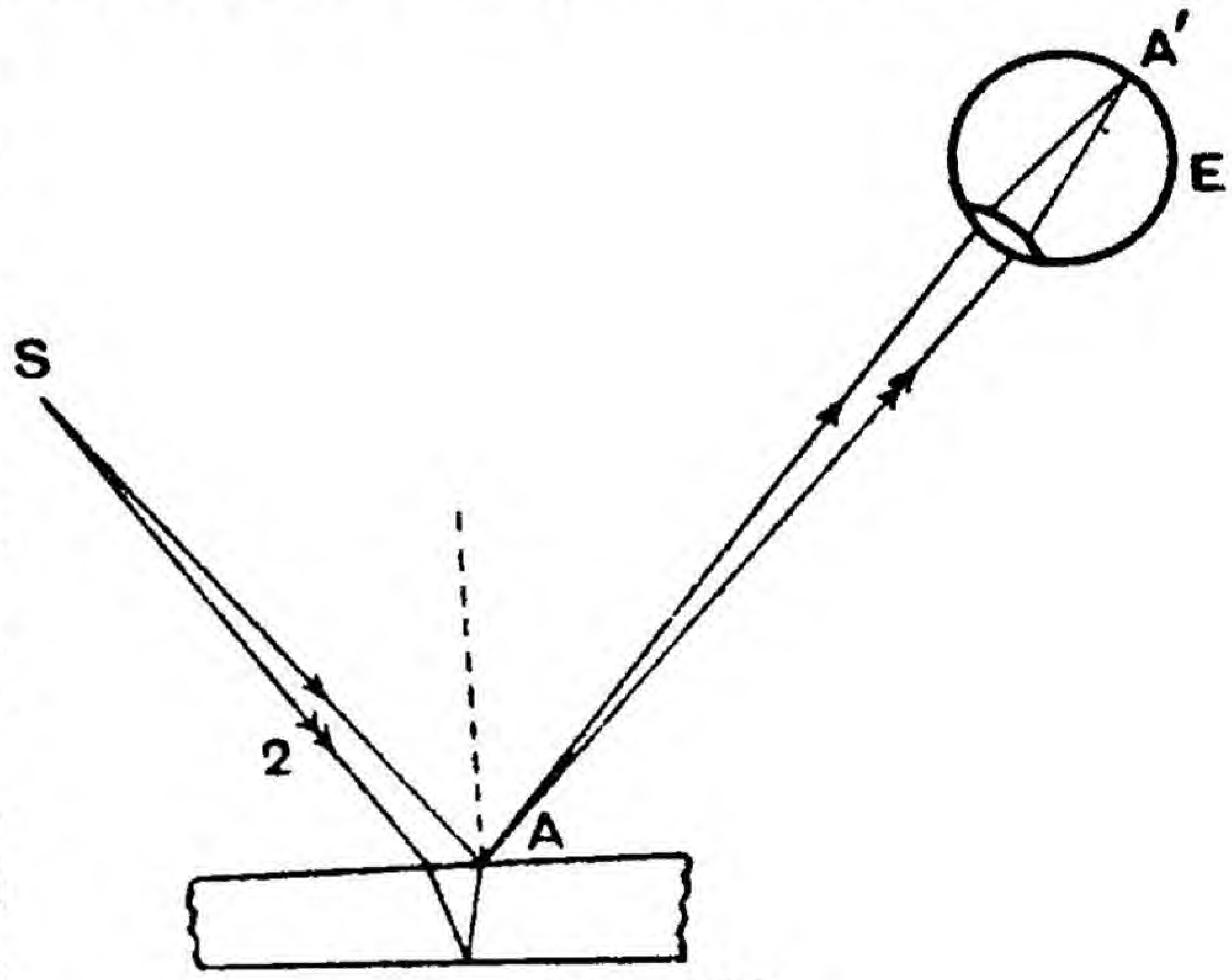


Fig. 203.

The path difference between the two rays can be calculated by the help of Fig. 204, in which it must be remembered that the film is only a few wave-lengths thick and that the angle θ between the two surfaces of the film is never more than a few minutes, while the source and eye are usually several centimetres away from it. Therefore the rays SA and SB are parallel within the limits of experimental error and, if BC is drawn perpendicular to SA, SB=SC. BEA is the path of the ray SB after striking the top surface of the film, AD is perpendicular to BE and BE is produced to F to make EF=EA. If AF is joined, it follows that EAF is an isosceles triangle, whose angle at E is bisected by the bottom surface of the film and so AF is equal to $2e$, where e is the thickness of

the film at the point A. Since the waves along SA and SB arrive at B and C simultaneously, the path difference between them at the point A is given by

$$\delta=(\text{BE}+\text{EA}) \text{ of film}-\text{AC of air}$$

But light goes from C to A in air, while it is going from B to D in the film;

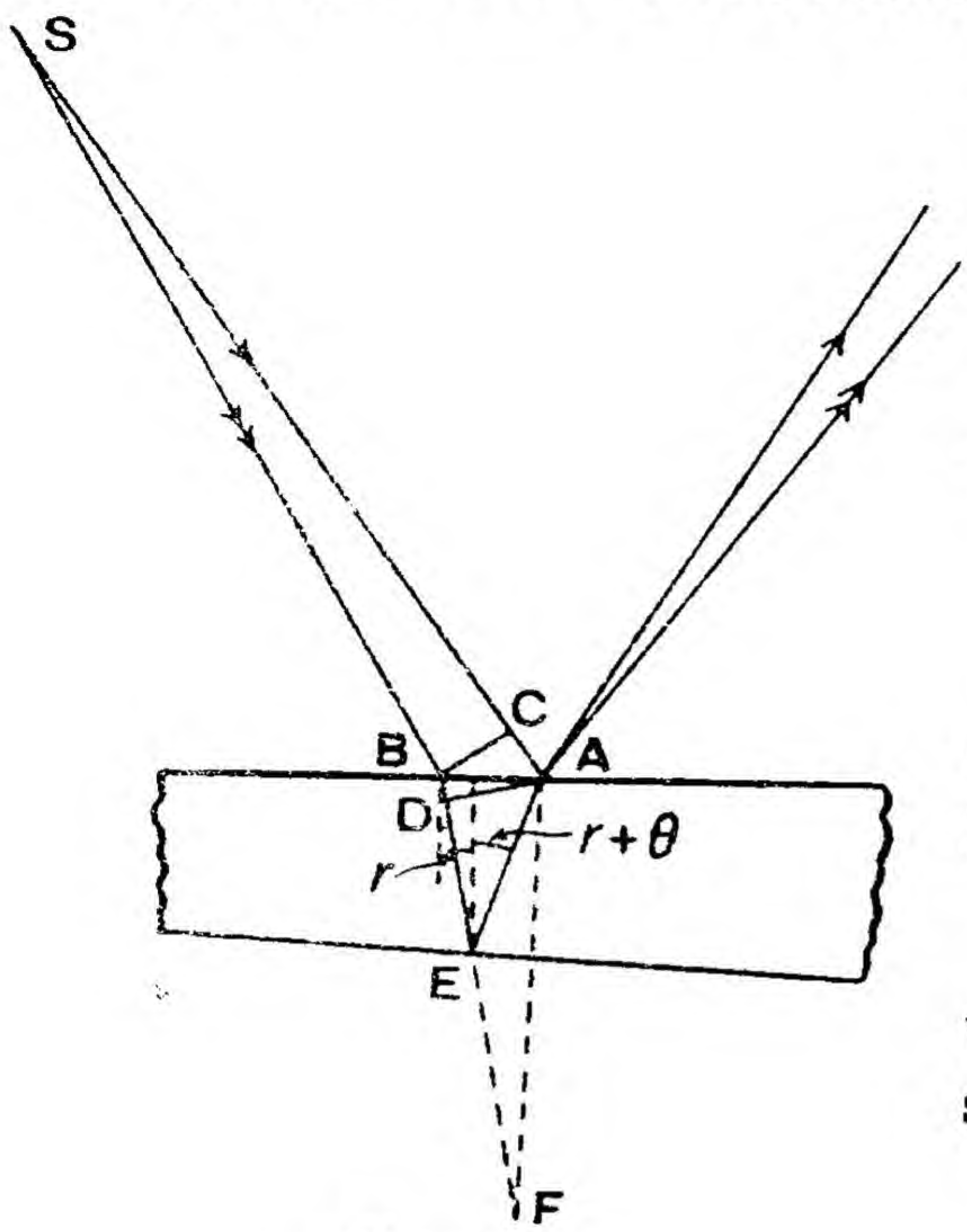


Fig. 204.

$$\therefore \text{ AC in air}=\text{BD in the film}$$

$$\therefore \delta=(\text{DE}+\text{EA}) \text{ of film}$$

$$=n(\text{DE}+\text{EA}) \text{ cm. of air}$$

$$=n\,\text{DF}$$

where n = the refractive index of the film.

$$\text{But DF}=\text{AF}\cos \text{EFA}$$

$$=\text{AF}\cos(r+\theta)$$

where r = angle of inclination of the ray SB in the film

$$=\text{AF}\cos r$$

within the limits of experimental error, since θ is only a few minutes of arc.

$$\therefore \delta=n\,\text{AF}\cos r$$

$$\therefore \delta=2ne\cos r \qquad (72)$$

We must emphasise that the rays SA and SB are practically parallel, as may be seen from the fact that AB is 10^{-4} cm. and SA $=10$ cm. in a typical case, so the angle r may be used indiscriminately to denote the angle of inclination in the film of either SA or SB.

We may now use this equation to predict what will happen in four specially simple cases. In the **first case** the incident light is parallel and monochromatic and the faces of the film are plane and parallel to one another. Both e and r are constant, so δ is constant for all points of the film. If it is an integral number of wave-lengths, the film will be dark all over, because of the π change in phase suffered by the ray SA at reflection at the more dense medium, there being no phase change when the ray SB is reflected at the less dense medium, air. If the path difference is an odd number of half waves, the illumination of the film is everywhere a maximum and, if it lies in between these possibilities, the illumination will be of constant intensity at every point. So in this case, **the illumination of the film seen by reflected light will be uniform.** In the **second case**, the incident light is parallel and white, the film being of constant thickness as before. Again e and r and therefore δ are constant, so those wave-lengths for which the path difference is an integral number of waves will be absent from the reflected light. **Hence the film will have a uniform coloration**; it will look the same colour all over. We should mention here that, if a broad beam of strictly parallel light is

incident on the film, only a small area AA′ (Fig. 205) of the film will be visible by reflected light at any one position of the eye; this is the area which can reflect light into the pupil of the eye and so will have about the same area as the pupil of the eye itself. Different parts of the film can be scanned by the eye, however, by moving it about and the point is that the intensity will remain the same with monochromatic light and the colour will remain the same with white light. In the **third case**, monochromatic parallel light is incident on a film of varying thickness, so r is constant, while e varies; hence δ varies too. There will be a minimum intensity in the reflected light for those thicknesses for which the path difference is an integral number of waves and a maximum intensity for thicknesses for which the path difference is an odd number of half waves. We shall see a set of alternate dark and bright bands, **any one band being a locus of constant path difference or constant thickness of the film.** The bands are like contours of the film. In the **fourth case**, if the incident light is white instead of monochromatic, we shall see a set of coloured bands, **any one band being a contour of the film.** This last case can be realised in the laboratory by sending a beam of parallel white light on to a vertical soap film formed on a circular wire frame and focussing an image of the soap film formed by reflected light on to a screen. As the soap solution drains to the bottom, making the film thinner at the top, a set of horizontal coloured bands forms, the colours becoming more brilliant as the film gets thinner. Finally a black band forms at the top of the film itself when its thickness is less than a wave-length, the blackness being due to the change in phase of the wave reflected at the front surface of the film.

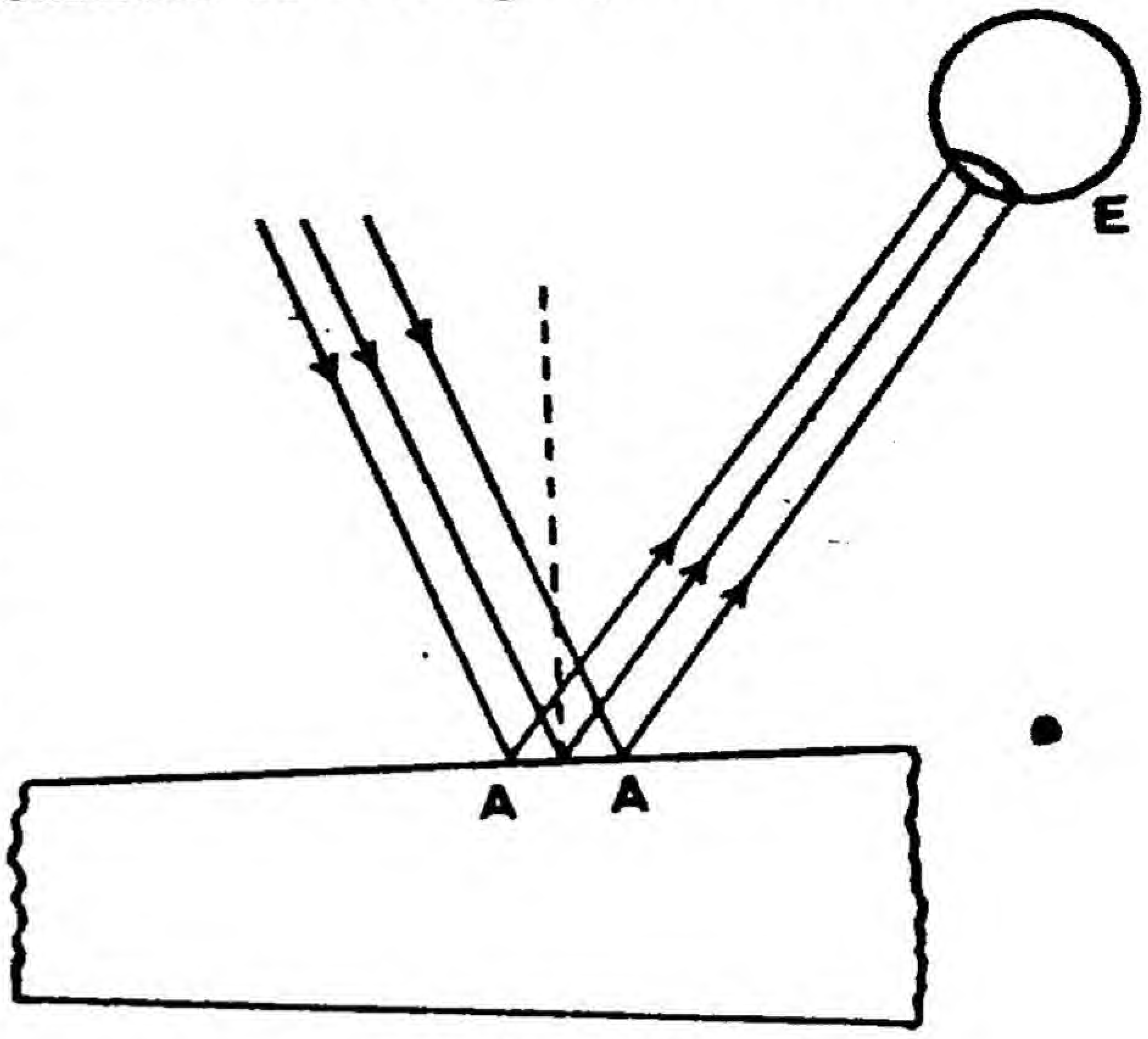


Fig. 205.

We can now consider the common case of the colours of a thin film of oil on a puddle of water seen by light reflected from the sky. It is clear from Fig. 206 that the point A of the film is seen by light which comes from the point S of the sky, while the point A′ of the film is seen by light coming from another point S′ of the sky. In general, each point of the film is seen by light from a different point of the sky, so that, *although an extended source is used, the condition of the interfering wave trains originating from a single point of the source is not violated.* S′ does send light on to A, but that light cannot enter the eye at E and contributes nothing to the illumination of the image of A formed by the eye, so that those rays can be ignored for our purpose. In this case, both e and r vary in the diagram. But, if we consider a practical case of a film

some 10 cm. in linear dimensions viewed by an eye 150 cm. vertically above the middle point of the film, for example, the least and greatest values

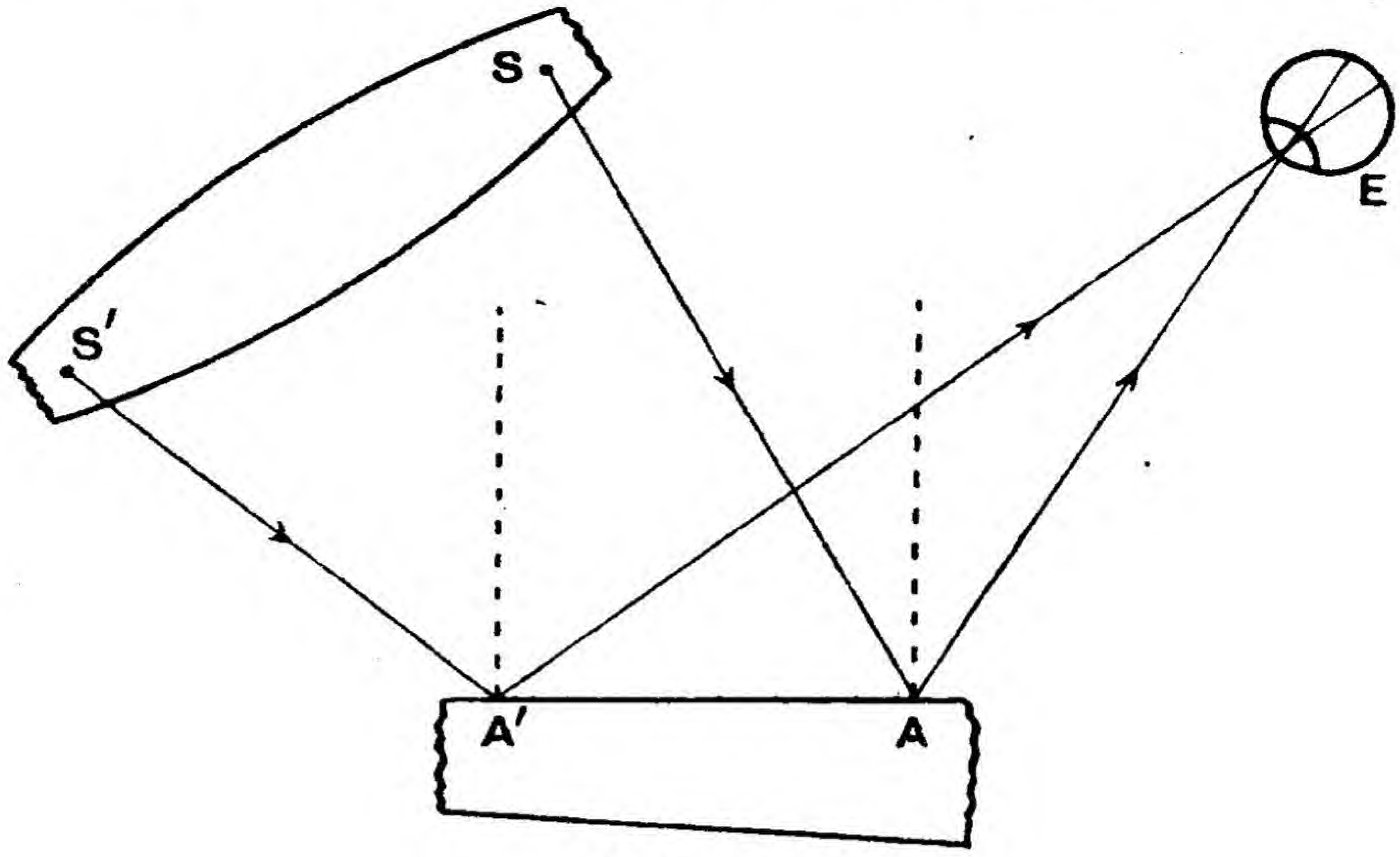


Fig. 206.

of r are 0° and 1·7°. The cosines of these angles are 1 and 0·9996, so the variation of path difference due to a variation in the angle of inclination of the light on the film can be ignored. **So we shall see a set of coloured bands, each band being a contour of equal thickness of the film.** For this reason, these bands are often called **fringes of equal thickness.** The reader can easily verify for himself that this is what he does see, and he will notice that the colour of any particular region of the film alters, if the head is moved, so as to change the angle of inclination of the light reaching his eye from that portion of the film. It will also be seen that a finite area of the film can be seen for one position of the eye with an extended source, which is the sole effect of using an extended source compared to a point source.

129. NEWTON'S RINGS

Newton's Rings can be seen and measured by the arrangement shown in Fig. 207. A source of monochromatic light such as a sodium flame, S, is placed in the focal plane of the collimating lens C, so that light from *one point* of the source emerges from the lens as a horizontal parallel beam and falls on the glass plate G at 45° to the horizontal, where some of it is reflected so as to fall normally on the lens L resting on the glass plate P. This lens is a converging lens of focal length about 100 cm. and the light reflected from the plate and this lens passes through the glass plate G into the microscope. If the microscope is focussed on the air film between the lens and plate, a set of alternate dark and bright rings is seen; the centre of the rings is dark and coincides with the point of contact of the lens and plate and the rings get closer and closer together as we go further out. These rings are known as Newton's Rings, as they were first observed by Newton, and we have already seen

(Art. 97) that he measured them with great care and suggested an explanation of them based on a combination of the corpuscular and wave theory, all the reflection occurring at the top surface of the plate.

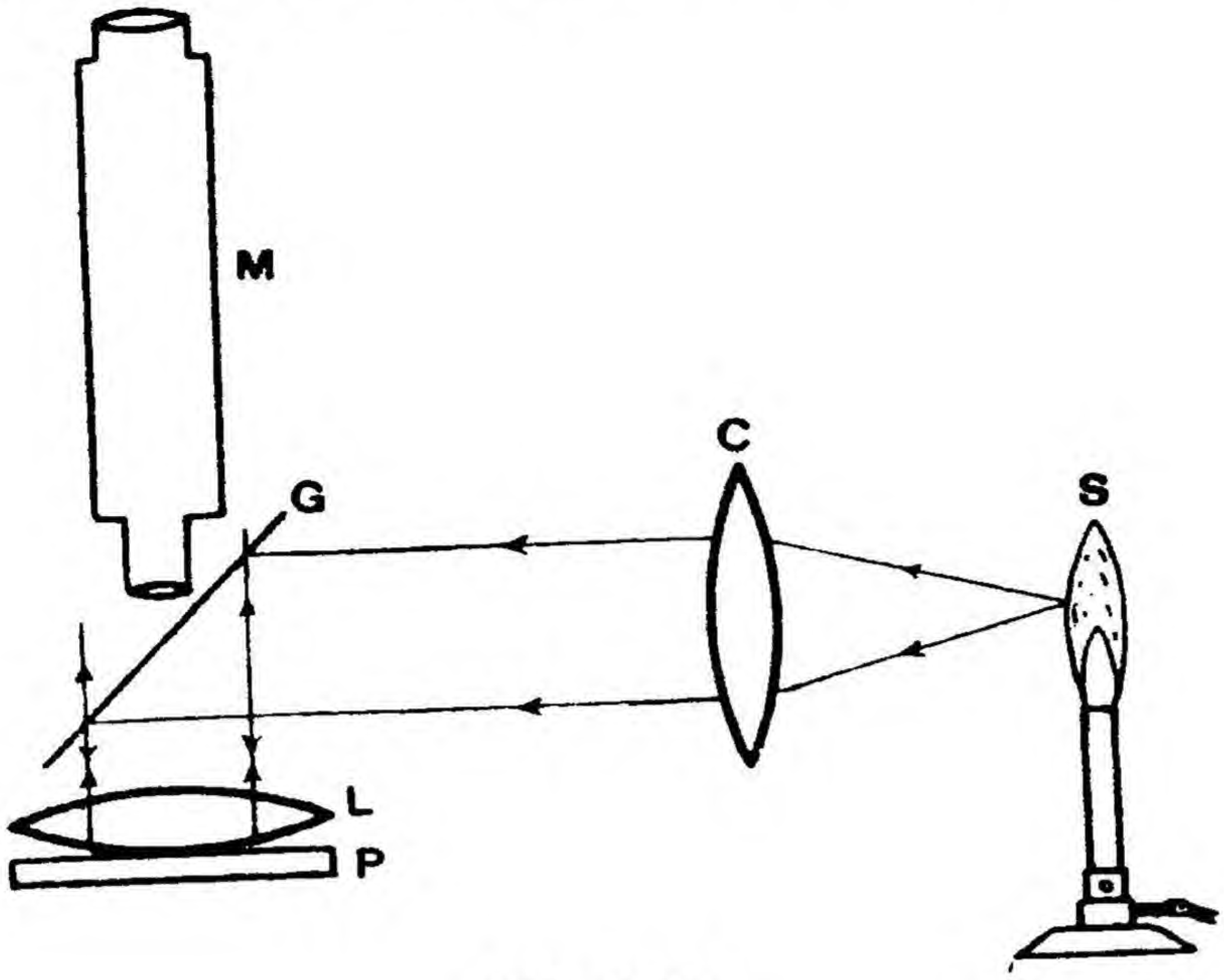


Fig. 207.

But we can now see that there is an alternative explanation of the rings due to interference between waves reflected from the bottom surface of the lens and the top surface of the plate ; in fact, they are the colours of the thin air film, lenticular in shape, enclosed between the lens and plate. Since the light falls normally on the air film, the path difference between the waves reflected from the two surfaces of the film, which take just the type of paths discussed in the previous article, is $2e$, taking the refractive index of air as 1. Since the thickness of the air film increases as we go outwards from the point of contact of the lens and plate, we shall get a dark ring at each thickness of the film for which

$$2e = n\lambda$$

where n is an integer. This is the condition for a minimum, not a maximum, since the light reflected from the top surface of the plate suffers a π change in phase, the other ray suffering no phase change. Each fringe is a locus of constant film thickness and so is a ring with the point of contact of the lens and plate as centre, since the bottom surface of the lens is a portion of the surface of a sphere. The rings are therefore another case of fringes of equal thickness.

The mathematical theory of the rings is as follows. If the n^{th} dark ring of radius ρ_n passes through A (Fig. 208), the path difference between the two interfering waves is 2 AB and is equal to $n\lambda$. Now AB=DE, and DE(2R−DE)=BD×DH, where R=radius of curvature of the bottom surface of the lens.

$$\therefore\ DE = \frac{BD^2}{2R}$$

since DE, which is of the order of 10^{-4} cm., is negligible compared to 2R, which is about 200 cm.

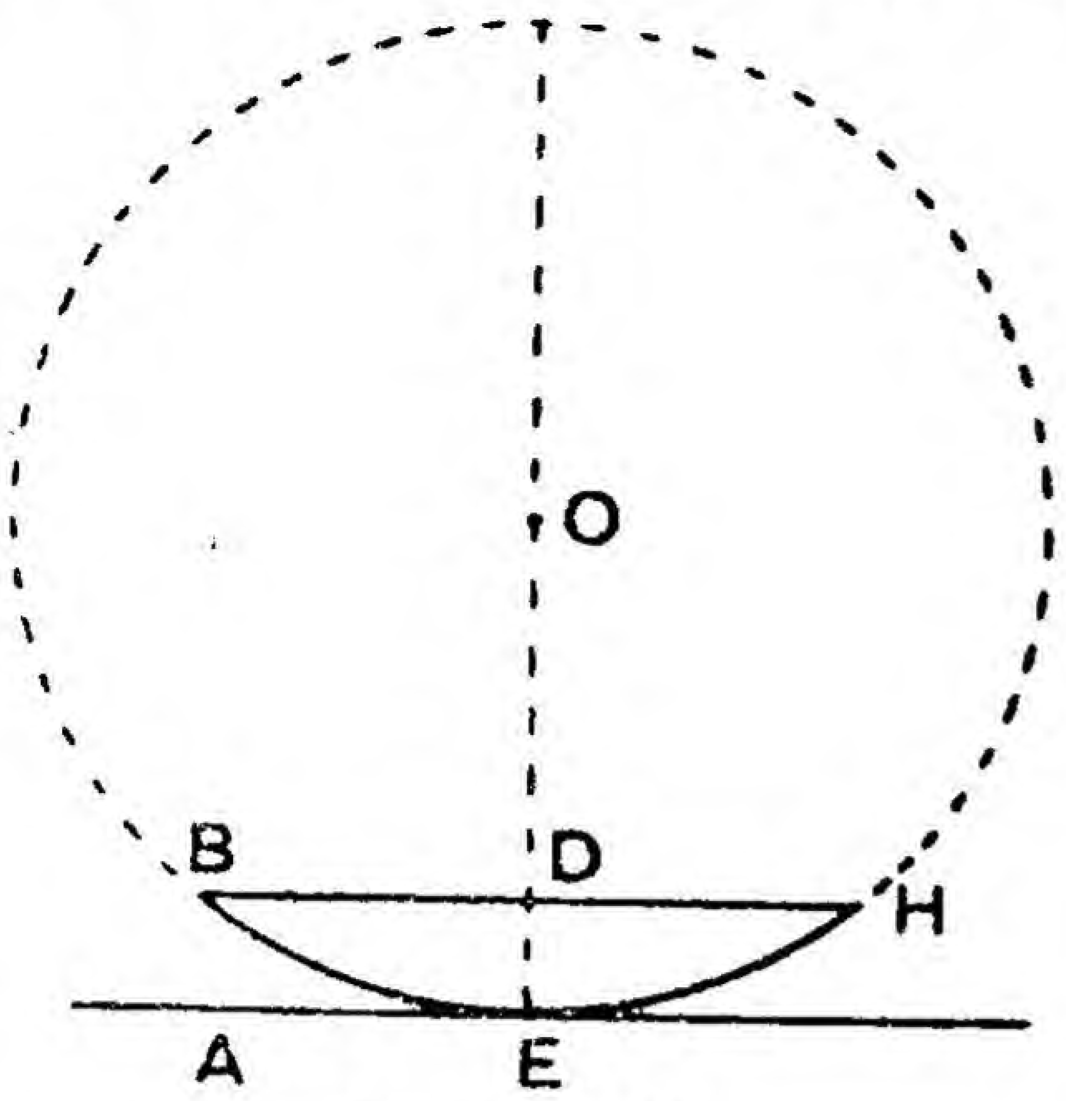


Fig. 208.

$$\text{But } BD = AE = \rho_n$$

$$\therefore DE = \frac{\rho_n^2}{2R}$$

$$\text{But } 2AB = n\lambda = 2DE$$

$$\therefore \rho_n^2 = \frac{n\lambda}{R}$$

$$\therefore \rho_n^2 = \lambda R n \quad \ldots\ldots\ldots (73)$$

$$\rho_n = \sqrt{n\lambda R} \quad \ldots\ldots\ldots (74)$$

Thus the radii of the rings are proportional to the square roots of the natural numbers, which Newton found from his measurements (Art. 97) and which explains why the rings get closer together as we go outwards.

The truth of this equation and of this explanation of Newton's Rings can be verified by measurements made to find the wave-length of light. The diameter of some twenty rings starting from the centre and working outwards is measured with a travelling microscope and a graph of ρ_n^2 against n is plotted and turns out to be a straight line. The slope of this graph is found and is $R\lambda$. The radius of curvature of the bottom surface of the lens is found either by a spherometer or preferably by an optical method, such as Boys' method, and from this and the slope of the graph the value of λ can be calculated. The result for sodium yellow light agrees with the values found by the previous methods within the limits of experimental error, thus confirming the truth of the wave theory explanation of Newton's Rings.

Which of the two explanations of Newton's Rings is the true one? Can we say that one is true and the other false? Is it not more nearly just to say that one is a closer approximation to the truth than the other? We should probably prefer the explanation given above, because it brings Newton's Rings into line with other phenomena as one of a number of cases of interference, and it is one of the functions of scientific investigation to show a connection between facts, which appear on superficial examination to be unrelated. Science is just ordered knowledge; it is the correlation of experimental facts into a rational scheme. And the wave theory explanation fits Newton's Rings into a scheme of facts related by the principle of interference. But a few further facts about Newton's Rings will settle the matter in favour of interference and against fits of reflection and transmission. We have seen that the interference explanation attributes the dark centre to the π change in phase of the wave reflected at the top surface of the glass plate. If a crown glass lens is placed on top of a flint glass plate with oil of sassafras between them, the crown glass is optically less dense than the oil, which is in turn less

dense than the flint glass. As the reflection of both waves takes place at a more dense medium, there will be a π change of phase in each case and the centre of the rings should be bright. Young tried this experiment and the centre was bright! It is difficult to see how Newton's theory of fits of reflection and transmission could lead to this result. Again, if the light is incident on the lens and plate at an angle r instead of normally, the path difference at a given point of the film decreases on the interference explanation, since it becomes $2e \cos r$, and so the diameter of the rings should increase, if the incidence is changed from normal to oblique. But on Newton's theory the diameter of the rings should decrease, since a corpuscle entering the film at a point of given thickness has to travel further between entering the air film and leaving it again, while the length travelled by a corpuscle between two fits of transmission is unaltered. Let a dark ring occur at a given point of the film with normal incidence due to a corpuscle passing through two fits of transmission between entering and leaving the film; if the incidence becomes oblique, the corpuscle will pass through more than two fits at that point of the film and so that dark ring will occur at a place of less thickness and its diameter will decrease. Experiment decides in favour of the interference once again, as the radii of the rings increase when the incidence becomes oblique. Finally, it can be shown that reflection from the lower surface of the lens plays an essential part in the formation of the rings by sending light at the polarising angle (Art. 148) on to the lens and replacing the plate by a mirror. The light reflected from the lens is then plane polarised and can be cut out by a suitable analyser, while that from the mirror is unpolarised and cannot be so cut out. According to Newton's theory, the rings should still be present, but according to the wave theory they should vanish. And they do vanish! So we are finally driven to accept the interference theory of Newton's Rings, partly because it does fit all the facts and partly because it is freer from *ad hoc* assumption. But, although we have now to reject Newton's explanation of Newton's Rings, there is much that we can learn from it. In the first place, we see how Newton did not hesitate to introduce an undulatory conception into hls theory of light, when the experimental facts seemed to warrant and indeed to demand it. Newton's genius possibly found its highest expression at this point in his work. Secondly, it is possibly this very introduction of the undulatory conception that may have led Young and others to go a step further and to try a more thoroughgoing application of wave theory to the problem. The reader will be making a big mistake if he regards Newton's explanation as useless, because it finally turned out to be wrong. The final truth about a thing only emerges after many years of toil and thought, after trial and error, after careful observation and the invention of hypotheses to fit the observations, and finally the formulation of a theory free of all inconsistency and capable of explaining the whole range of facts under investigation. Newton's theory played its part in the discovery of

ultimate truth and that is why it has been mentioned here. Newton himself said : " If I have seen further than others, it is because I have stood on the shoulders of giants "—a charming tribute of a great man to the work of those who had gone before him. And what better tribute could be paid to the man himself, for Young would probably have agreed that Newton's work set him thinking about waves in connection with Newton's Rings ?

130. FURTHER CONSIDERATIONS

We shall now consider a few points of interest about interference in thin films, which may conveniently be taken at this stage. The reader will notice a striking difference between these fringes and the ones previously considered. They were non-localised, while Newton's Rings, for example, cannot be seen unless the microscope is focussed on the film between the lens and plate. This is because the two wave trains responsible for the interference effects at a given point cross only in the neighbourhood of the film, as is evident from Fig. 203, and the reader will recall that interference is only seen at the place where the two wave trains cross. So this class of interference fringes is called **localised fringes.**

A rather simple and instructive example of interference in thin films is provided by a wedged-shaped film of air enclosed between two glass plates, illuminated in just the same way as the lens and plate in Newton's Rings. If the angle of the wedge is a few minutes, the reader should be able to prove that a set of equally spaced alternate dark and bright bands parallel to the edge is seen in reflected light. These bands are another case of fringes of equal thickness. If the angle of the wedge is α, the fringe width is $\frac{\alpha\lambda}{2}$.

If Newton's Rings are examined by transmitted light by placing the microscope below the lens and plate, a set of rings complementary to those seen by reflected light is produced. The centre is bright and, to a dark ring by reflected light, there corresponds a bright ring by transmitted light. But there is one significant difference between the two sets of rings. Those by transmitted light are much poorer in contrast than those in reflected light. The reason for this is that the amplitude of the ray transmitted directly through the air film is so much greater than that which suffers two internal reflections in

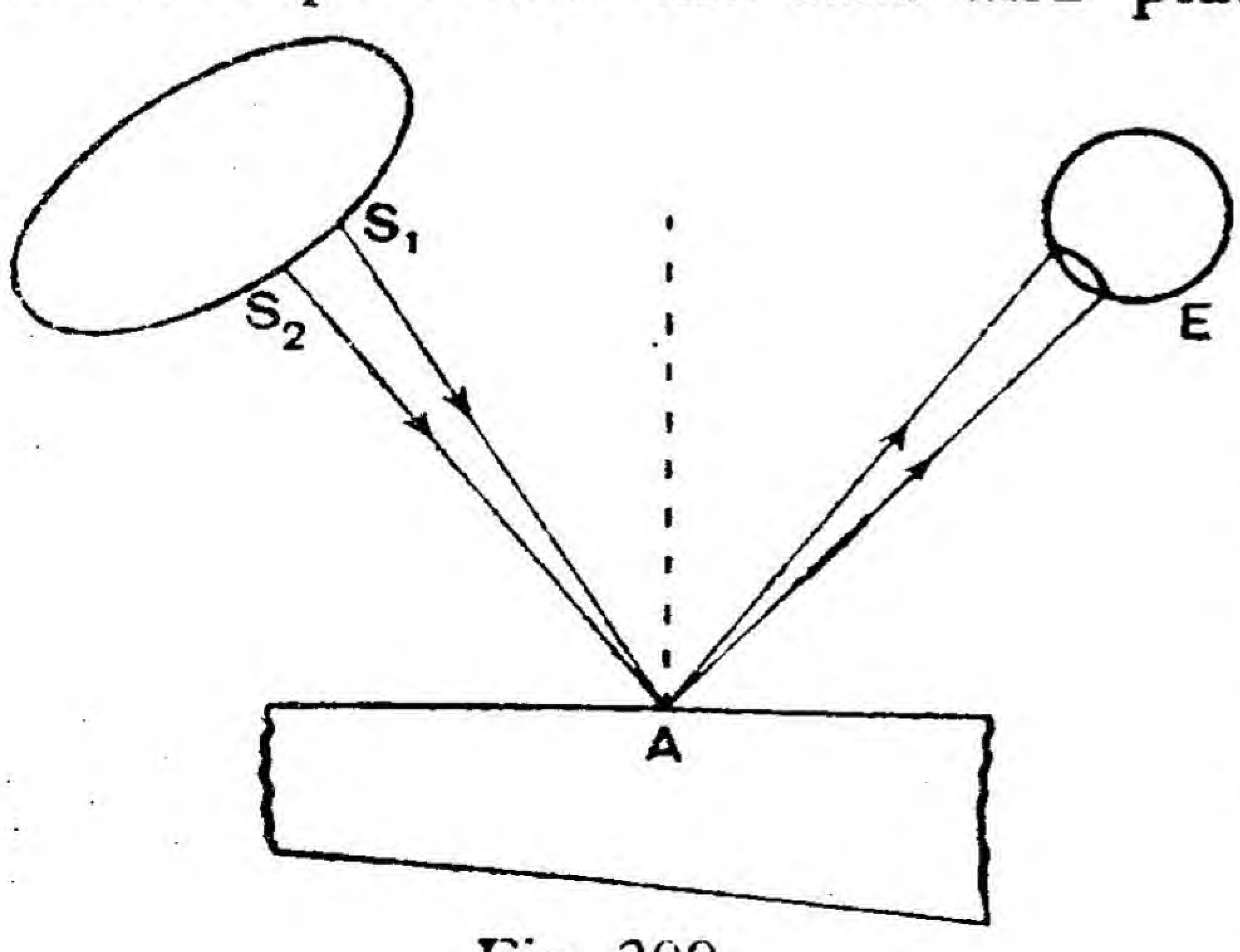


Fig. 209.

the air film that there is not much difference between the amplitudes of the maxima and minima. The colours of all thin films can be seen by transmitted light and the same considerations apply.

We have already proved that many more rings can be observed in monochromatic light than in white light using Newton's Rings apparatus, and we have seen the reason for it. But even with monochromatic light the rings eventually disappear. Why cannot interference be obtained with a thick film? Why is a thin film necessary? It arises from the fact that the pupil of the eye has a finite diameter and, if an extended source is used, the eye can receive light from the region S_1S_2 of the source in focussing on the point A of the film (Fig. 209). The angle of inclination of the ray from S_1 is a little different from that from S_2 and so the path difference of the pair of rays from S_1 is a little different from that of the pair from S_2. The difference can be found by differentiating $2e \cos r$ and is $2e \sin r dr$. This variation in path difference gets greater, the greater e, and when e becomes so big that it amounts to half a wave-length no interference can be seen. For, if the two waves derived from the ray from S_1 destroy each other at A, the pair from S_2 will reinforce and so the effect of the beam of rays accepted by the eye will always be the same, in other words, interference effects have ceased. So we should only expect to get them with thin films.

131. FRINGES OF EQUAL INCLINATION

We have finally to consider a case of interference which has been used to replace the metre by the wave-length of light as a standard of length. It is the case of the thin film in which the thickness is constant while the angle of inclination in the film varies enough to produce an alteration in the path difference of several wave-lengths. A horizontal cross-section of the apparatus used for obtaining the fringes seen under these conditions is sketched in Fig. 210. A pin-hole S, illuminated by monochromatic

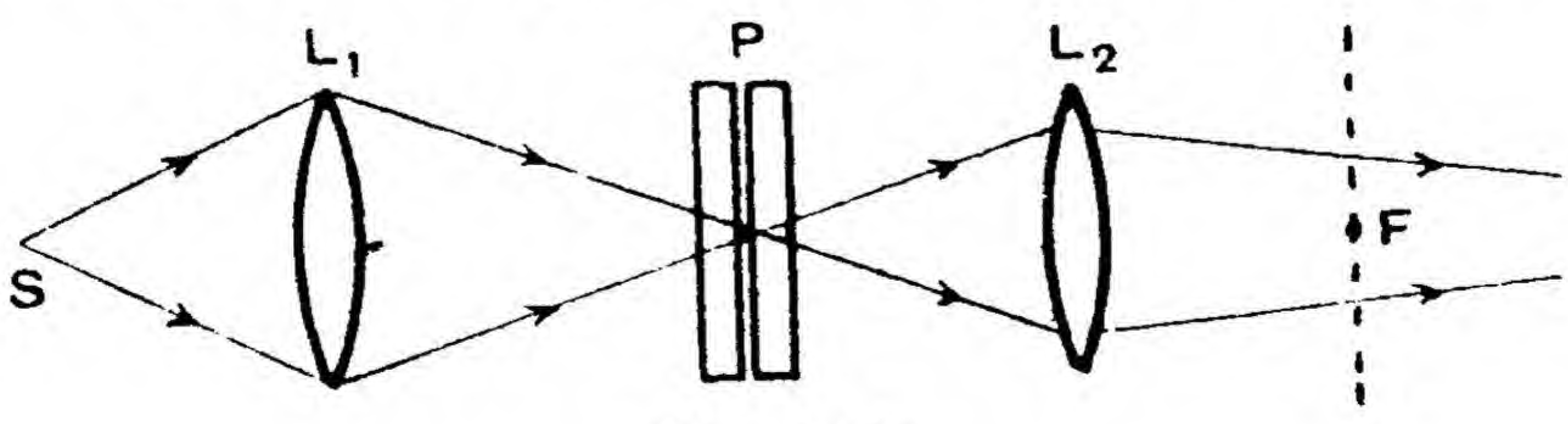


Fig. 210.

light, sends light on to a lens L_1 which converges it to a thin film of air contained between two half-silvered plates P, whose faces are optically plane and parallel. The term half-silvered means that the whole area of the plate is covered with silver to such a density that half of the incident light is reflected and the other half transmitted. The term optically plane means that the surface of the plate is a plane to within a tenth of the wave-length of the light and the plates must be parallel to

the same degree of accuracy. The light then diverges on to the lens L_2, each ray being split up, a part of it going straight through the air film, the rest being reflected twice, four times, six times, and so on before finally emerging from the film. Circular fringes, whose centre coincides with the focus of the lens L_2, can be seen in the focal plane of that lens. They can be magnified by means of an eyepiece placed beyond the lens L_2, if this is desired.

The fringes are produced in the following way. Any particular ray falling on the air film at an angle r (Fig. 211) will suffer a number of internal reflections in passing through the film, so that it is split up into a number of parallel rays, one ray having passed straight through the film, the others having suffered two, four, six . . . reflections. It can be proved in a similar way to that used in Art. 128 that the path difference between any two successive rays at the line AB at right angles to them is $2e \cos r$, the refractive index of air being 1. These parallel rays are brought to a focus at the point F_r in the focal plane of the lens L_2 and they will have the same path difference at the point of crossing F_r, as they did at the line AB. There will be a maximum or minimum

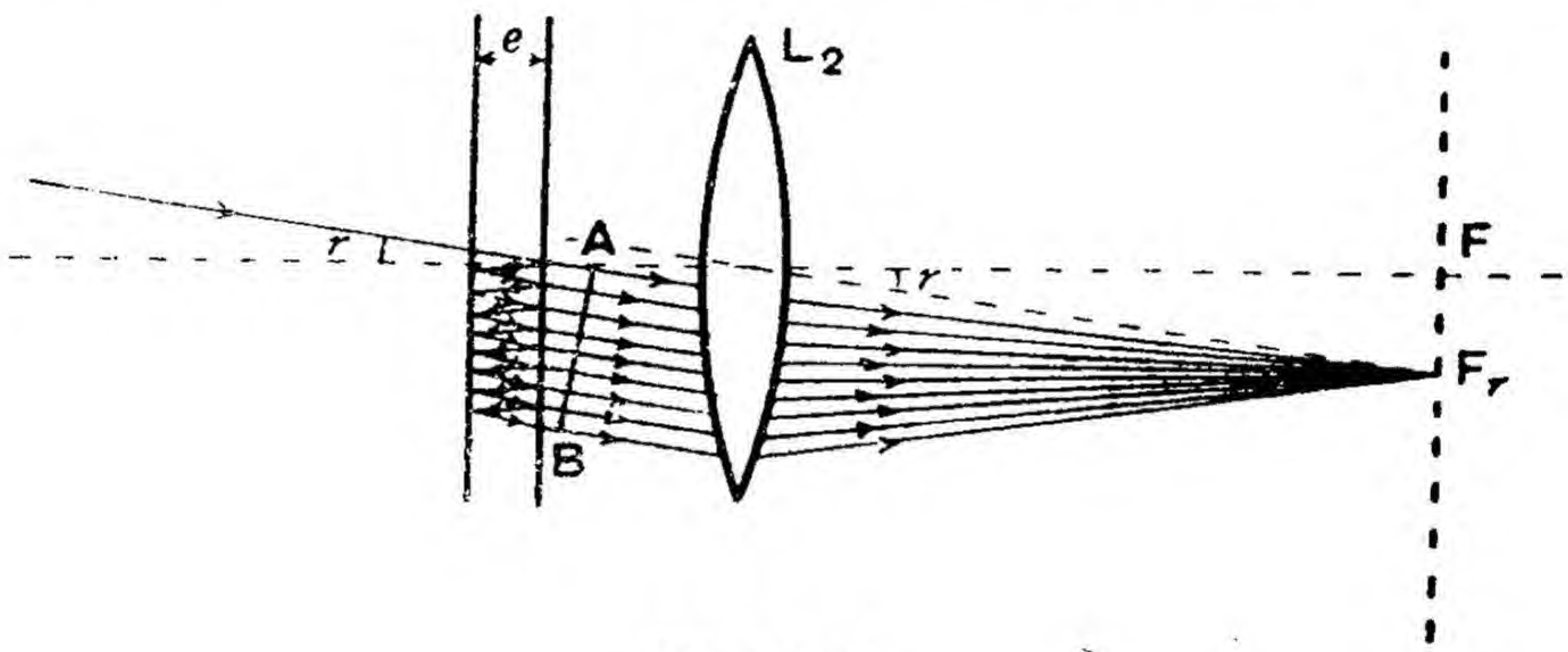


Fig. 211.

at F_r according as $2e \cos r$ is $n\lambda$ or $(2n+1)\frac{\lambda}{2}$. Let us imagine that F_r is situated at a maximum. Then, since the actual apparatus is symmetrical about the common axis of the two lenses as axis, the diagram can be rotated about that axis and the point F_r describes a circle with F as centre and with radius subtending an angle r at the centre of the lens L_2. We shall see a set of bright rings whose radii subtend an angle r at the centre of the lens L_2, where r is given by

$$2e \cos r = n\lambda,$$

n being an integer. The innermost ring with the *least* value of r is the ring with the *greatest* value of n, the value of n corresponding to each ring decreasing as we go outwards. This is just the *opposite* to Newton's Rings, where the order of interference, as n is sometimes called, increases as we go outwards. But the two sets of rings are similar in that they get closer together as we go outwards. Each ring is a locus of rays falling on the film at the same angle and so they are called **fringes of equal inclination,**

which should be contrasted with Newton's Rings, which are fringes of equal thickness. The fringes are only seen where the interfering waves cross, which is in the focal plane of the lens L_2. But this lens is really producing an image of the actual fringes themselves, which are therefore localised at infinity behind the film. They can be seen with the naked eye held close up to the film, being located at infinity behind the film at the place where the parallel rays emerging at any particular angle meet when produced backwards.

These fringes can be used to measure the thickness of the air film. The parallel plates P (Fig. 210) are mounted on a spectrometer turntable and a pin is mounted so that its tip coincides with the focus of the lens L_2. The fringes are obtained and the turntable is rotated in a horizontal plane until the centre of the first bright ring coincides with the tip of the pin. The reading of the turntable is recorded and it is turned round until the centre of the other side of the same ring coincides with the tip of the pin. The reading of the turntable is taken and the difference between these two readings is $2r$, the angle subtended by the first ring at the centre of the lens. Similar readings are taken for as many rings as can be conveniently and accurately measured. Then we have

$$2e \cos r = n\lambda$$

for the first ring,

$$2e \cos r_1 = (n-1)\lambda$$

for the second ring,

$$2e \cos r_s = (n-s)\lambda$$

for the $(s+1)^{th}$ ring. Subtracting the third from the first equation, we have

$$2e(\cos r - \cos r_s) = s\lambda$$

from which e can be calculated.

The importance of this method lies in the fact that these fringes can be obtained with films 10 cm. thick, and the thickness of such a film can then be obtained in terms of the wave-length of a suitable monochromatic light to an accuracy of less than 1 in 1,000,000. This 10-cm. film can then be compared with the standard metre by a step-up method, enabling the wave-length of light to be used as a standard of length instead of a bar of platinum. This is obviously desirable, as the properties of an atom are likely to be more constant than those of a bar of platinum, which is an artificial thing. The reader must consult more advanced text-books on Optics for details of this important determination, as what has been written above is only intended to indicate the lines along which the problem is attacked.

132. RAYLEIGH REFRACTOMETER

The Rayleigh Refractometer, which is a simple modification of Young's Slits, can be used for the measurement of the refractive index of gases and the angular diameter of giant stars. It consists of two parallel slits

A and B mounted before the objective L of a telescope (Fig. 212) illuminated by parallel monochromatic light from a point source S. If the axis of the telescope is parallel to the direction of the incident light, the waves from A and B which go straight on will be in phase when leaving A and B and will reach the focus F of the objective at the same time and produce a bright band there. This will coincide with the vertical cross-wire of the telescope. But the waves which are diffracted through an angle θ will have a path difference BC at the line AC normal to the direction in which they are travelling, so that they will reach their focus F_1 with that path difference. Now, as $BC = e \sin \theta$, we shall get a series of maxima in directions θ given by the equation

$$e \sin \theta = n\lambda$$

where n is an integer. For example, if the case shown in Fig. 212 represents the first such maximum, then $e \sin \theta = \lambda$ and the rays in that direction come to a focus at F_1, the point in the focal plane of the objective cut by the ray through the middle of the lens making an angle θ with its

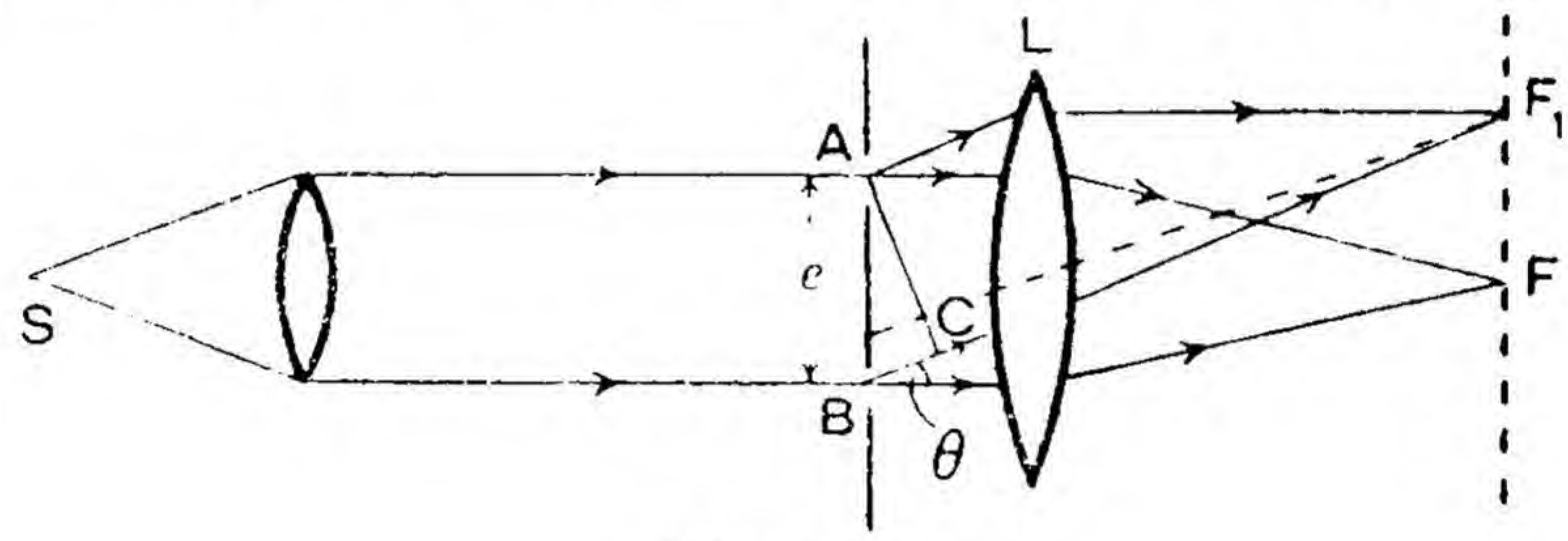


Fig. 212.

axis. There will be similar maxima at corresponding points in the focal plane of the objective separated by minima. These fringes are really the equivalent of Young's fringes and they are present on each side of the focus of the objective. If two similar tubes are now placed in the path of the rays entering the slits A and B and the air is gradually removed from the one in front of A, the fringes will slowly move downwards across the vertical cross-wire of the objective as the air is exhausted. This shows that air has a refractive index greater than 1 relative to a vacuum, since the downwards movement of the central fringe means that a single crest originating from S now arrives at A before B, and the optical path of the ray going to A has been shortened. If s fringes pass the vertical cross-wire of the objective, as the pressure p of the air in the tube of length l is reduced to 0, then a length l of air has been replaced by a length l of vacuum in the ray entering A. If n is the refractive index of air at pressure p, the optical length of the ray entering A has been changed from nl to l, that is, it has been shortened by $(n-1)l$. But each time the path is shortened by λ, one fringe passes the cross-wire. Since s fringes pass in this case

$$(n-1)l = s\lambda$$

from which n can be calculated, when s has been counted.

We can see how the angular diameter of a star can be measured by considering first of all the case of two stars subtending an angle θ at the objective of the telescope. If the axis of the telescope be pointed at one of the stars, we know that the Rayleigh Refractometer will produce fringes with a maximum at the vertical cross-wire, for the star is a distant point source and two rays from it to the slits A and B will be parallel to the axis of the telescope. But two rays from the other star to A and B will be parallel to one another at an angle θ to the axis of the telescope (Fig. 213). Consider rays from this star which are diffracted from A and B in a direction parallel to the axis of the telescope and come to a focus at F. The two waves are in phase at B and C, so their path difference at the point F

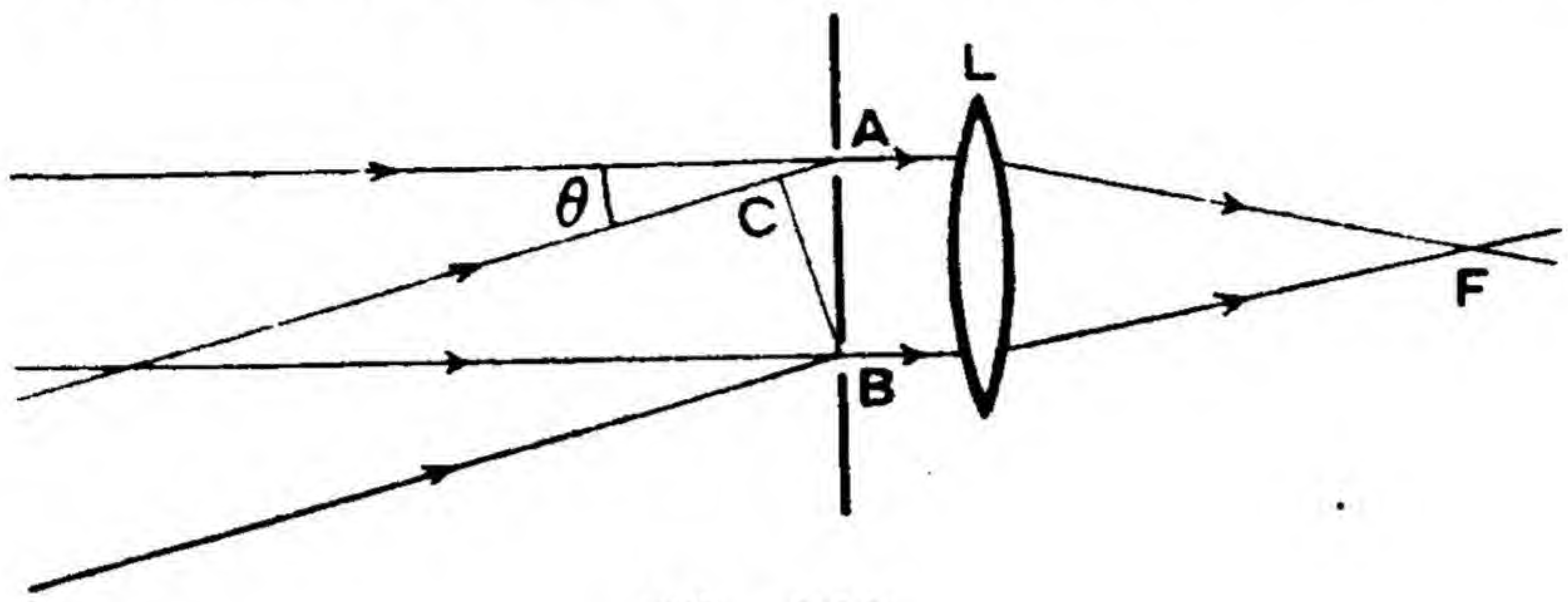


Fig. 213.

where they cross is AC, since the time taken by the light to go from A or B to F is the same. But AC$=e\sin\theta$, and if this is $\frac{\lambda}{2}$, there will be a minimum at F. So this star will produce a fringe system with a minimum at the vertical cross-wire, and its fringe system and that of the other star will produce uniform illumination, since the minima of one coincide with the maxima of the other and vice versa. Hence the fringe system disappears. If two stars subtending an unknown angle θ at the earth send light on to a Rayleigh Refractometer and the distance apart of the slits is adjusted to the value e needed to make the fringes disappear, the maxima of one system coincide with the minima of the other ; in other words, $e\sin\theta=\frac{\lambda}{2}$ and θ can be calculated. If we now point the Rayleigh Refractometer at a single star, we can analyse it into pairs of points, such as a point at the centre and a point on the circumference of the star. We can treat each pair of points as two separate stars and the fringes due to them will disappear if the distance apart of the slits, e, is adjusted so that

$$e\sin\frac{\theta}{2}=\frac{\lambda}{2}$$

or

$$\theta=\frac{\lambda}{e}$$

as θ is so small. If this condition is satisfied for one pair of points, it will be satisfied for all of them and the single star will produce no fringes. The reader will recognise that this is similar to the disappearance of the

fringes with Young's Slits if the single slit S is made too wide (Art. 125). So the distance apart of the slits in the refractometer is adjusted for disappearance of the fringes and the angular diameter of the star is calculated from the above equation. The experiment was first carried out by Michelson and the distance e was 306·5 cm. for Betelgeuse, a giant red star in the constellation Orion. Taking λ as $5{\cdot}75 \times 10^{-5}$ cm. this gives an angular diameter of 0·047″, from which the linear diameter turns out to be 240 million miles !

We have travelled a long way since we started on the road labelled interference, not quite knowing what we should find at the end of it. What have we found ? We have found that interference is possible with light indeed it was there all the time, if only we had had the wisdom to see it. We have found that the wave-length of light is very small, which is just what we need to explain away rectilinear propagation, and we are just beginning to realise that these very small light waves may come in very useful for measuring lengths accurately. They have been used to measure the coefficient of expansion of crystals (Barton's "Heat," Ch. 3) ; they can be used to test the flatness of surfaces ; they can be used to measure the refractive indices of gases, which are much too small to be found directly, and the angular diameters of stars ; finally they bid fair to oust the standard metre as the standard of length. It is the old story again, that new knowledge is always revolutionising industry ; knowledge found for its own sake is often useful in practical affairs. But we cannot travel further along this road and we turn, confident of success, to our next problem : diffraction !

EXAMPLES ON CHAPTER XIII

1. Under suitable conditions two beams of light may combine to produce darkness over limited regions. What are the necessary conditions which must be satisfied if interference fringes are to be produced ? Illustrate your answer by reference to a simple interferometer such as Fresnel's mirrors.

The interfering sources in a simple interferometer are situated d cm. apart and the fringes produced are observed on a screen D cm. away. Find an expression for the distance between the centres of successive fringes in terms of d and D, and hence show that this distance w is given by

$$w = \frac{\lambda}{\phi}$$

where ϕ is the angle subtended at the screen by the two sources and λ is the wave-length of the light used. *(Camb. Schol.)*

2. Explain what is meant by the interference of light.

Describe in detail an interference experiment which could be used to measure the wave-length of sodium light, and make an estimate of the order of accuracy of this determination. Draw a diagram of the apparatus and explain in detail how the eye sees the interference fringes. In the apparatus you describe, what is the effect of substituting white light for the sodium light ? *(Camb. Schol.)*

3. Describe Fresnel's bi-prism method of obtaining interference fringes. In an experiment with this apparatus sodium light is used and bands 0·0196 cm. in width are observed at a distance of 100 cm. from the slit. A convex lens is then put between the observer and the prism so as to give an image of the "sources" at a distance 100 cm. from the slit. The distance apart of the images is found to be 0·70 cm., the lens being 30 cm. from the slit. Calculate the wave-length of the sodium light. (*Camb. Schol.*)

4. Under what conditions can interference of light take place? Describe some simple experiment based on interference for the measurement of the wave-length of light. (*O. and C.*)

5. Under what conditions can two sources of wave motion produce interference fringes? Describe an experimental arrangement for observing and measuring interference fringes in the case of light and show how the wave-length can be deduced.

Mention one case of interference in wave motion other than light. (*O. and C.*)

6. Describe in detail an experiment with Fresnel's biprism to determine the wave-length of light from a suitable source, proving the requisite formulæ.

Why is it necessary in such an experiment to use a narrow slit, whereas Newton's rings may be observed with light from an extended source? (*London B.Sc.*)

7. Discuss with examples the conditions necessary for the production of interference fringes in a region where two beams of light overlap.

The inclined faces of a glass bi-prism ($\mu = 1{\cdot}5$) make angles of 2° with the base of the prism. The slit is 10 cm. from the prism, and is illuminated by light of wave-length 5900 A.U. Calculate the spacing of the fringes observed at a distance of 1 metre from the prism. (*London B.Sc.*)

8. Describe and explain (a) one experiment which suggests that light is propagated by means of waves, (b) another experiment which shows that the wave-length of red light is greater than that of blue light. (*N.U.J.B.*)

9. A pair of fine parallel slits 0·35 mm. apart held close to the eye produces coloured interference fringes when a straight electric lamp filament is viewed at a distance of 6 metres. Explain this phenomenon.

Suggest a simple method of measuring the wave-length of light with this arrangement without the addition of any lenses or mirrors. Estimate the expected accuracy. (*N.U.J.B.*)

10. Describe the interference fringes produced by Fresnel's inclined mirrors. It is required to produce fringes of width 0·1 mm. in sodium light at a distance of 40 cm. from the source with this arrangement. Calculate the angle at which the two mirrors must be inclined to one another if the wave-length of sodium light is $5{\cdot}9 \times 10^{-5}$ cm. and the source is 6·0 cm. from the line of intersection of the mirrors.

11. Describe the production of interference fringes by Lloyd's single mirror. In what way do these fringes differ from those produced by Fresnel's bi-prism, for example? How can achromatic fringes be produced?

12. What are the conditions for the interference of light? How would you measure the radius of curvature of a plano-convex lens of very large radius of curvature? (*Oxford Schol.*)

13. A wedge of air of angle 0·4° is formed by two half-silvered glass surfaces, and is illuminated by sodium light. Draw rough scale diagrams showing the separations of the fringes which are formed near the apex of the wedge and at a distance of about 2 cm. from the apex. (Wave-length of sodium lines = 5890 and 5896×10^{-8} cm.) (*Camb. Schol.*)

14. Describe how (a) parallel and (b) circular interference fringes can be obtained using two plane glass plates. What is the effect of partially silvering the plates?

Discuss carefully the question of the localisation of the fringes in the two cases, and illustrate diagrammatically a simple arrangement by means of which the circular fringes might be photographed. (*Camb. Schol.*)

15. How do you account for the colours of thin films? Describe experiments which support your explanation, and show how the thickness of the film may be estimated.
(*Oxford Schol.*)

16. Give an account of the formation of the colours of thin films. A thin film of air is contained between two parallel-faced glass plates, the distance and angle between which can be varied. How would you adjust the plates in order to obtain (a) straight, (b) circular, interference fringes? Describe, with the help of diagrams, the methods of illumination and of viewing the fringes you would use.
(*Camb. Schol.*)

17. Newton's Rings are formed with sodium light ($\lambda = 5{\cdot}9 \times 10^{-5}$ cm.) between a plane glass plate and a convex lens surface. The diameters of two successive dark rings are 2·0 mm. and 2·236 mm. What is the radius of curvature of the lens surface?
(*Camb. Schol.*)

18. Describe how you would use Newton's Rings to measure the wave-length of light and give an outline of the necessary theory.

Newton's Rings are formed by reflection in the air film between a plane surface and a spherical surface of radius 50 cm., and it is noticed that the centre of the system is bright. What do you conclude from the fact that the centre is bright? If the diameter of the third bright ring is 0·181 cm. and the diameter of the twenty-third bright ring is 0·501 cm., what is the wave-length of the light used?
(*Camb. Schol.*)

19. Explain the formation of Newton's Rings between a convex lens and a plane glass plate for monochromatic light.

Describe the alteration of the fringe system when a liquid having a refractive index intermediate between that of the lens and that of the plate is introduced into the space between them.

Newton's Rings are formed with reflected light of wave-length 5890×10^{-8} cm. using a plano-convex lens and a plane glass plate with liquid between them. The diameter of the third bright ring is 2 mm. If the radius of curvature of the curved surface of the lens is 90 cm., find the refractive index of the liquid.
(*Camb. Schol.*)

20. Explain the colours of thin films. (*O. and C.*)

21. Explain the formation of Newton's Rings. A drop of oil of refractive index 1·58 is placed between a plano-convex lens and a plane glass plate on which the former rests. The lens and plate are of the same refractive index, 1·50. Calculate the diameter of the fifth bright ring if the radius of curvature of the lens surface is 200 cm., and light of wave-length 5900×10^{-8} cm. is reflected normally from the system. What effect will be observed when the plate is replaced by another of refractive index 1·65?
(*London B.Sc.*)

22. Explain the formation of Newton's Rings.

Newton's Rings are observed between a plane surface and a lens supported at a variable distance above it. Sodium light is used and it is found that for certain distances the ring system disappears. Explain this phenomenon and calculate the distances of separation for which the fringes would be invisible. Explain the fact that for all distances greater than a certain value the fringes are invisible. (Wave-length of D_1 line of sodium is 5896 Å.U. Wave-length of the D_2 line is 5890 Å.U.)
(*Tripos, Part I.*)

23. Newton's Rings can be seen in the light transmitted through the lens and plate. State and explain any differences between these rings and those seen in reflected light.

24. Fringes are produced by Young's Slits with light of wave-length $7{\cdot}5 \times 10^{-5}$ cm. and the slit of a spectrometer is placed to coincide with the tenth bright fringe. White light is now substituted for the red light. Describe precisely and exactly the appearance of the spectrum.

25. Two thin plates of glass separated by a distance of $5 \cdot 10^{-3}$ cm. are half-silvered on their inner surfaces. A beam of white light passes through them at right angles and falls on the slit of a prism spectrometer. Describe and account

for the appearance of the spectrum observed. Suggest a practical use for this arrangement. (Half-silvered surfaces reflect half and transmit half the light falling on them.) *(Oxford Schol.)*

26. Give an account of the interference fringes which may be formed by a wedge-shaped film of air enclosed between two flat glass plates. Describe an experiment for the production of these fringes.

If fringes of this kind are produced by white light, what will be seen in a spectroscope viewing them if its slit is set parallel to the edge of the wedge? *(Camb. Schol.)*

27. Explain the " colours of thin plates."

White light is incident on two parallel glass plates separated by an air film 0·001 cm. in thickness, and the reflected light is examined by a spectroscope. Find the number of dark bands seen in the spectrum between wave-lengths 4 and 7×10^{-5} cm., when the light is incident at an angle of 30° to the normal to the surfaces. *(London B.Sc.)*

28. How would you attempt to measure the refractive index of hydrogen, given that the refractive index of air at N.T.P. is 1·0003? *(O. and C.)*

29. Describe a method of measuring the refractive index of a gas.

In an experiment with a Jamin interferometer the length of the gas-filled tube was 20 cm. On changing the pressure of the gas by 70 cm. of mercury, 70 fringes passed the cross wire of the observing instrument. Find the value of the refractive index of the gas at a pressure of 76 cm. of mercury, and at the temperature at which the experiment was carried out. *(Tripos, Part I.)*

30. Show how the angular diameter of a star can be measured by an interference method.

Chapter XIV

DIFFRACTION

133. INTRODUCTORY

We have obtained interference and have learned from it that the wave-length of light is very small, of the order of 5×10^{-5} cm. From what we know of water and sound waves, it is not surprising that we have not noticed diffraction with light, for diffraction in water and sound waves gets less as the width of the obstacle or aperture increases compared with the wave-length of the waves. But we must now proceed a stage further than mere argument by analogy and make use of our theory, or model, of waves to do two things. We must deduce from it why light does travel so nearly in straight lines in ordinary practice. Then we must use it to tell us how to set up experiments which will demonstrate beyond all doubt the diffraction of light. Here we see the power of a scientific theory, in that it enables us to discuss and consider cases in which the conditions are widely different from anything which we can demonstrate for certain in the laboratory. For example, if we are right about the wave-length of light, the obstacles used in ordinary experiments on rectilinear propagation of light are some 1,000,000 times bigger than the wave-length of the waves. It would be most awkward to do similar experiments with water waves to see what to expect in such cases. The theory saves us from the necessity of doing the experiment. But the theory is really more important than that; it is the working thought model which the scientific man makes of the known world; it is his artistic creation and perhaps it is more real to him than the world itself. It is as far on one side of the practical everyday world as poetry is on the other. Let us now see how our theory of waves "explains away" rectilinear propagation.

134. RECTILINEAR PROPAGATION

Let us consider the propagation of a plane wave represented by WW and in particular let us calculate its effect at the point P (Fig. 214). We have to do this assuming the truth of Huygens' principle, so we treat each point of the wave front as the source of secondary wavelets and our problem is to add up their amplitudes at the point P. One difficulty confronts us at once; wavelets from different parts of the wave front

are in general at different distances from P and so will arrive at P with different phases, a crest from one point with a trough from another and intermediate phases from other points. The point P, in fact, has a vibration due to each wavelet of the same period but different phases and probably different amplitudes. How are we to sort out this confusion? One way would be to write down the equations of the vibrations of P due to each wavelet and to add them together, but there is a simpler way due to Fresnel which gives quite accurate results. He divides up the wave front into zones so that, at P, the wavelets from one zone are just π out of phase with those from the next. When a crest arrives at P from one zone, troughs are arriving from the next, and it is easy to add up crests and troughs. This division into **half-period zones** is done in the following way. PO is drawn normal to the wave front to cut it at O, O being called the **pole** of the wave front relative to the point P. With centre P, draw a set of spheres of radii $d+\frac{\lambda}{2}$, $d+\frac{2\lambda}{2}$, $d+\frac{3\lambda}{2}$, $d+\frac{4\lambda}{2}$ and so on, d being the distance PO. These spheres cut the wave front in a set of circles of radii OA, OB, OC, OD, and so on. They divide it into a set of half-period zones, the first zone being the circle of radius OA,

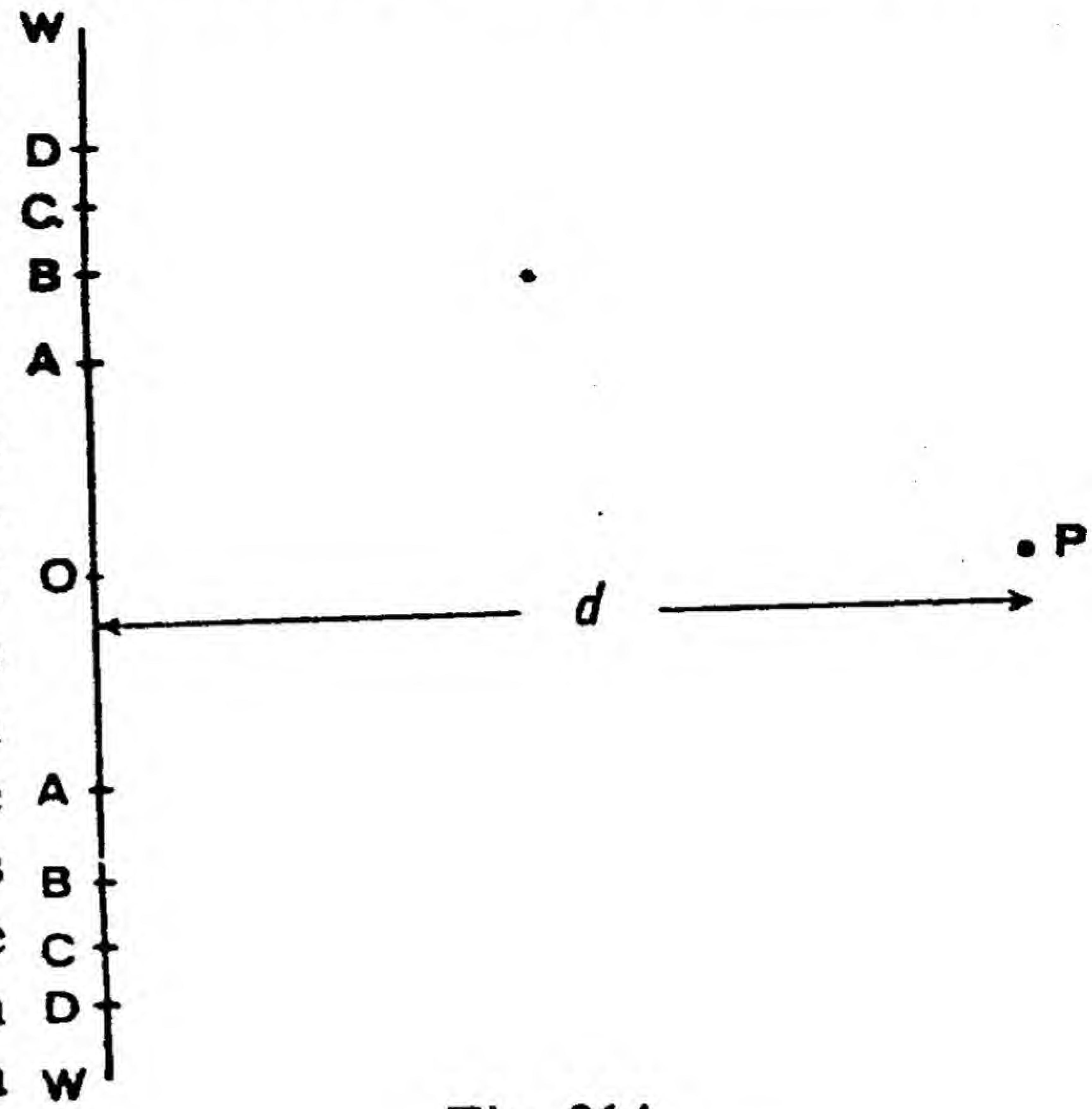


Fig. 214.

the second the annulus between the circles of radii OA and OB, and so on. Since BP is $d+\frac{2\lambda}{2}$ and AP is $d+\frac{\lambda}{2}$, any crest from the circumference of the first zone arrives at P with a trough from the circumference of the second zone. And what is true of these particular parts of the two zones is also true of any parts. If we consider the region on the circle R_1 (Fig. 215) in the first zone, there is always a corresponding circle R_2 in the second zone, such that crests from R_1 arrive at P at the same time as troughs from R_2. Hence, if we call the amplitude of all the secondary wavelets from the first zone at P, $+s_1$, we may call that of the second zone $-s_2$, since

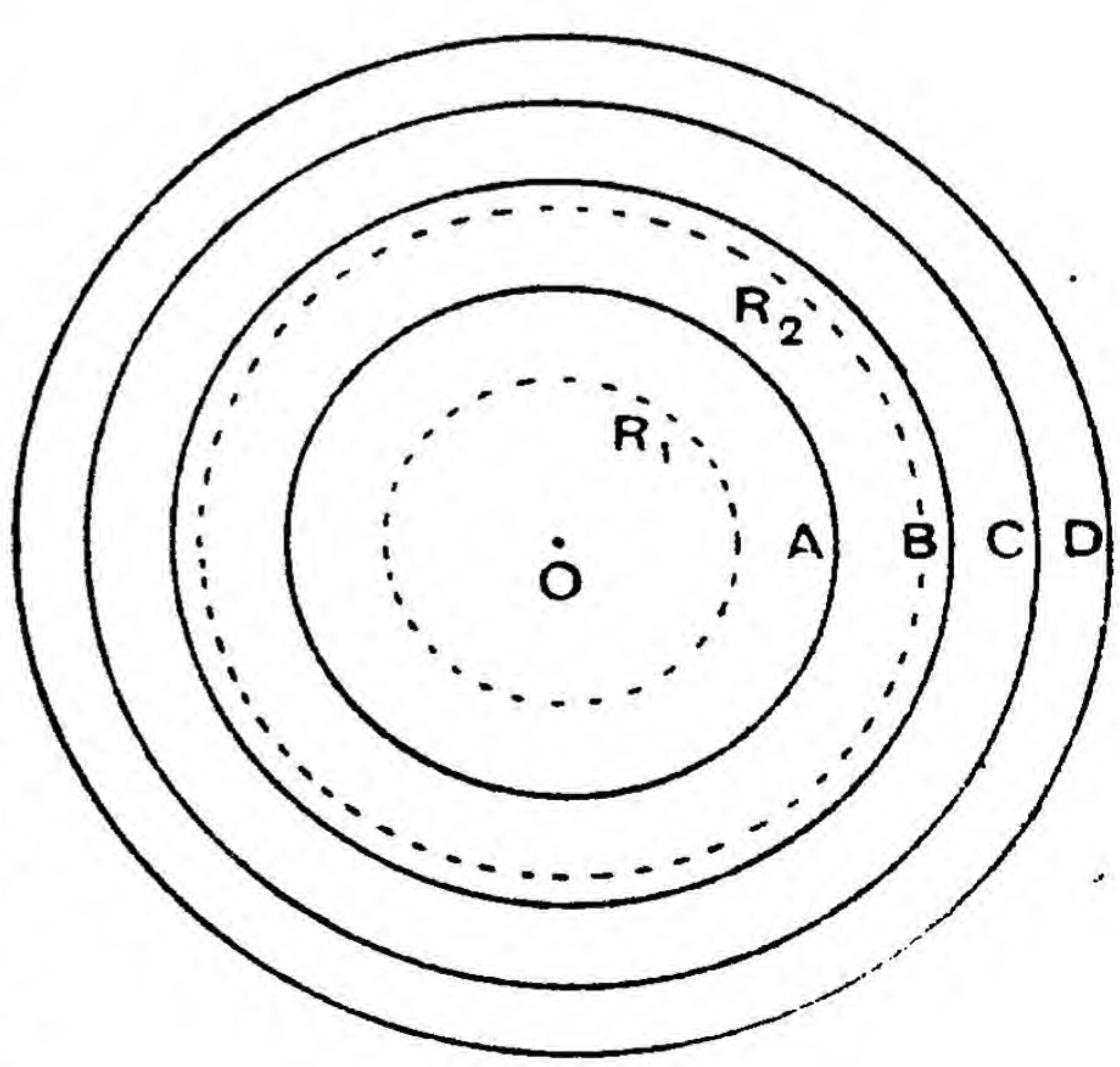


Fig. 215.

they are π out of phase with those from the first zone. Similarly we may call the amplitudes at P of the wavelets from the 3rd, 4th, 5th . . . zones, $+s_3$, $-s_4$, $+s_5$. . . . How will these amplitudes vary as we go from zone to zone? They will depend on the areas of the zones, their distances from P, and their obliquities to the line OP. The radius, r_n, of the nth circle is given by

$$r_n{}^2=\left(d+\frac{n\lambda}{2}\right)^2-d^2$$

$$\therefore\ r_n{}^2=nd\lambda$$

since $\frac{n^2\lambda^2}{4}$ is negligible compared to $nd\lambda$, when λ is of the order of 10^{-5} cm.

Therefore the radii of the zones are proportional to the square roots of the natural numbers, and the zones get closer together as we go outwards. The area of the n^{th} zone, which is contained between the n^{th} and $(n-1)^{th}$ circles is given by

$$\begin{aligned}\text{Area}&=\pi r_n{}^2-\pi r_{n-1}^2\\&=\pi nd\lambda-\pi(n-1)d\lambda\\&=\pi d\lambda\end{aligned}$$

So the areas of the zones are constant and this causes no diminution in amplitude of the wavelets at P. But the zones get further away from P and the wavelets fall on P more obliquely to OP as we go outwards and it is only the resolved part of the wavelet along OP that we want. Both of these factors cause the amplitude to decrease as we go outwards, so s_1, s_2, s_3, s_4, s_5, form a set of numbers of slowly decreasing magnitude. Hence the total amplitude S of the whole wave front at P is given by

$$S=s_1-s_2+s_3-s_4+\ \ldots\ +s_n,$$

where n is the total number of zones in the wave front. Let n be odd. We may arrange the series in two ways

$$S=\frac{s_1}{2}+\left(\frac{s_1}{2}-s_2+\frac{s_3}{2}\right)+\left(\frac{s_3}{2}-s_4+\frac{s_5}{2}\right)+\ \ldots\ +\left(\frac{s_{n-2}}{2}-s_{n-1}+\frac{s_n}{2}\right)+\frac{s_n}{2}$$

or $$S=s_1-\frac{s_2}{2}-\left(\frac{s_2}{2}-s_3+\frac{s_4}{2}\right)-\left(\frac{s_4}{2}-s_5+\frac{s_5}{2}\right)-\ \ldots$$

$$-\left(\frac{s_{n-3}}{2}-s_{n-2}+\frac{s_{n-1}}{2}\right)-\frac{s_{n-1}}{2}+s_n$$

Every expression in brackets is the arithmetic mean of the amplitudes of two alternate zones minus the amplitude of the intermediate zone. There are three possibilities about these expressions; they may be all positive, or all negative, or some may be positive and some negative. Consider the first possibility. If it is true,

$$\frac{s_1}{2}+\frac{s_n}{2}<S<s_1-\frac{s_2}{2}-\frac{s_{n-1}}{2}+s_n$$

If the second alternative is true,

$$s_1-\frac{s_2}{2}-\frac{s_{n-1}}{2}+s_n<S<\frac{s_1}{2}+\frac{s_n}{2}$$

In each case S lies between the same two limits and, if the third alternative is true, it is still nearer to these two limits. *So S lies definitely between those two limits* and, if the set of numbers s_1, s_2, s_3 . . . decreases very slowly in magnitude, the two limits coincide in the following expression for S.

$$S=\frac{s_1}{2}+\frac{s_n}{2} \qquad \text{(75)}$$

Let us work out the radius of the first zone for a typical experiment to demonstrate rectilinear propagation, in which a plane wave falls on an obstacle which casts a shadow on a screen some 20 cm. away. In this case the distance d is 20 cm. and taking λ as 5×10^{-5} cm., we get $r_1=\sqrt{20\times5\times10^{-5}}=3{\cdot}16\times10^{-2}$ cm., so that the diameter of the first zone is less than a millimetre and that of the ten thousandth zone is less than 10 cm., which is quite a normal width for the wave front falling on the obstacle. Now the amplitude of this 10,000th or n^{th} zone at P is quite negligible, and so the amplitude of the whole wave front at P reduces to half the amplitude of the first half-period zone, that is, of the portion of the wave front within about 0·3 mm. from O, the pole of P. To put it in another way, it is only the light starting from the first zone which is effective at P, the remainder of the wave front merely destroying half the effect of this zone by destructive interference. So light travels from a region around O, 0·3 mm. in radius, to P; in fact, it travels just about in the straight line OP.

We can see in another way how rectilinear propagation follows from this theory by considering a plane wave represented by WW (Fig. 216) falling on an obstacle T such as a sheet of cardboard, the shadow being cast on a screen some 20 cm. beyond the obstacle. The illumination at a point such as P_1, whose pole O_1 is so far from the edge of the obstacle that the first 100 or so zones pass it, will be just the same as if the obstacle were removed, for we may fairly assume that the zones after the hundredth add very little to the amplitude at P_1.

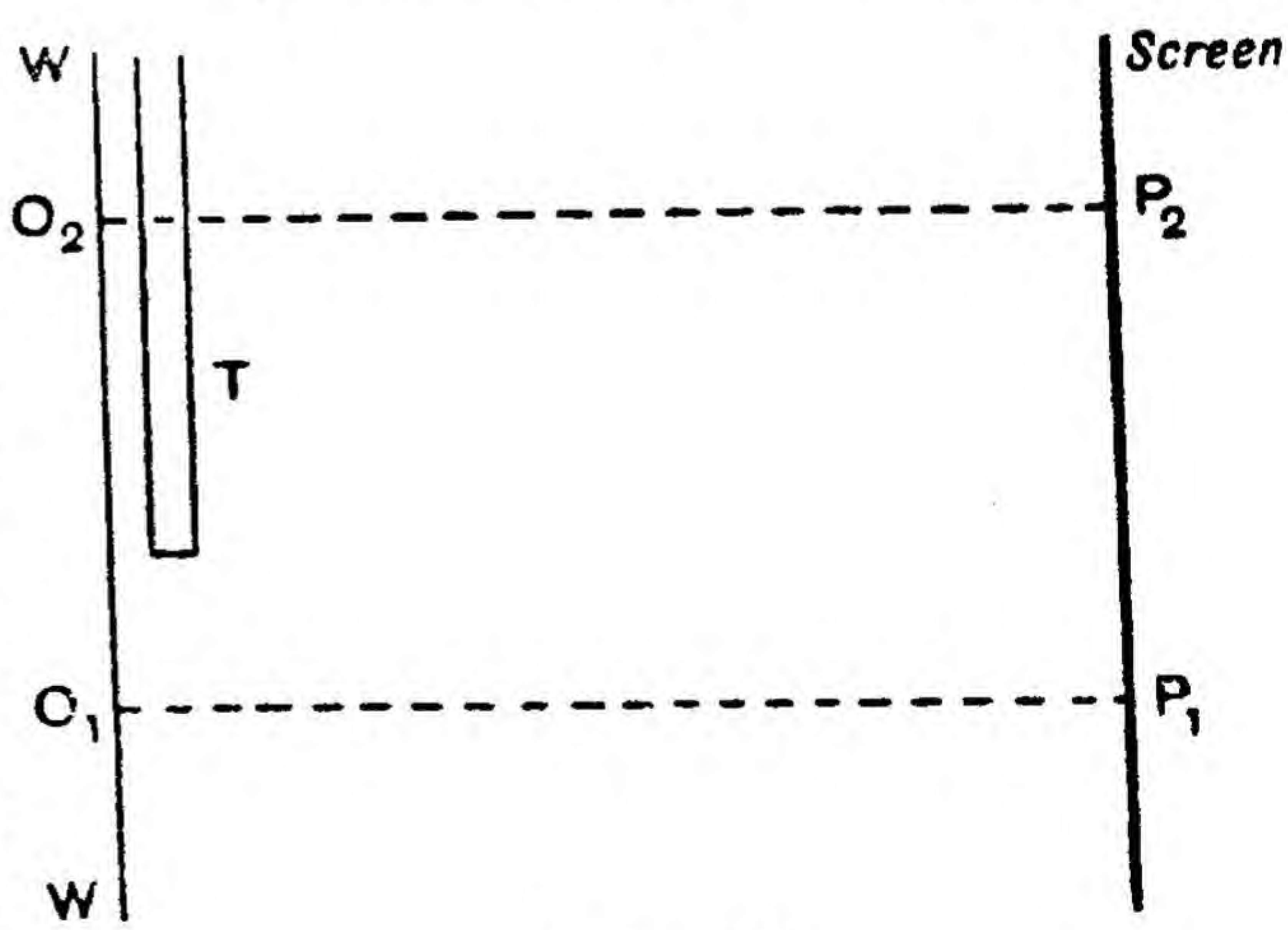


Fig. 216.

So we shall not notice the presence of the obstacle while O_1 is further from its edge than r_{100}, which is 3 mm., that is, while P_1 is further from the edge of the geometrical shadow by the same distance. In the same way, if P_2 is so far inside the geometrical shadow that its pole O_2 is further from the

edge of the obstacle than 3 mm., only zones of higher order than the hundredth can send wavelets to P_2. The shadow will then be quite dark, just as dark as if light travelled strictly in straight lines. It is only when O_2 is within 3 mm. of the edge of the obstacle and P within 3 mm. of the edge of the geometrical shadow that we should expect any anomalous effects. The choice of 100 zones as being the total number effective will be justified later (Art. 135). We have already seen (Art. 2) that it is impossible to produce a shadow with a perfectly sharp edge and we could not decide if it was due to the impossibility of getting a geometrical point image or to the fact that light does not travel strictly in straight lines. At any rate, we see that the very small wave-length of light will only produce deviations from rectilinear propagation within a millimetre or two of the edge of the shadow in a typical experiment, so it is no wonder that it has never been detected, because it is just in this region that the penumbra due to the finite size of the source occurs.

135. THE REALISATION OF DIFFRACTION

This way of explaining away rectilinear propagation leads at once to an experiment to demonstrate diffraction in light. Suppose that we put a circular obstacle between the wave front WW (Fig. 214) and the point P, arranging it so that its centre is on the line OP just to the right of O. It will cover up the first r zones, so that the amplitude of the wave front at P will be given by

$$S=\frac{S_{r+1}}{2}$$

If r is small enough there should be a spot of hght at the centre P of the shadow of the obstacle. But if we are to cut out only the first ten zones, for example, the obstacle must be only $\sqrt{10\times 20\times 5\times 10^{-5}}$ or 1 mm. in radius, if OP is 20 cm. It is hard to make an obstacle strictly circular when it is as small as this and the above reasoning may break down if there is a small departure from strict circular shape, as the zones are so small. Is our idea doomed to failure ? This is just where the theory comes to our aid, for it tells us that the number of zones cut out can be reduced either by decreasing the radius of the obstacle or *by increasing the size of the zones by making the distance d bigger*. If the distance is increased from 20 cm. to 200 cm., then the radius of the tenth zone becomes $\sqrt{10\times 200\times 5\times 10^{-5}}$ or 0·316 cm. and it is quite feasible to make a strictly circular object of this size. When Poisson presented this result to the Paris Academy of Sciences, that there is a bright spot of light at the centre of the shadow of a circular object cast by a point source of light, his conclusion was held to be so absurd that it disproved the wave theory by the method of *reductio ad absurdum* ! But Poisson had other ideas and he asked his friend Arago to try the experiment, which can be done in the following way. S is a pin-hole illuminated by an arc lamp (Fig. 217) and C is a

threepenny-piece of 8 mm. radius some 2 metres away. As the waves from S are practically plane waves by the time they reach the obstacle, the foregoing theory holds. If the shadow of the threepenny-piece is

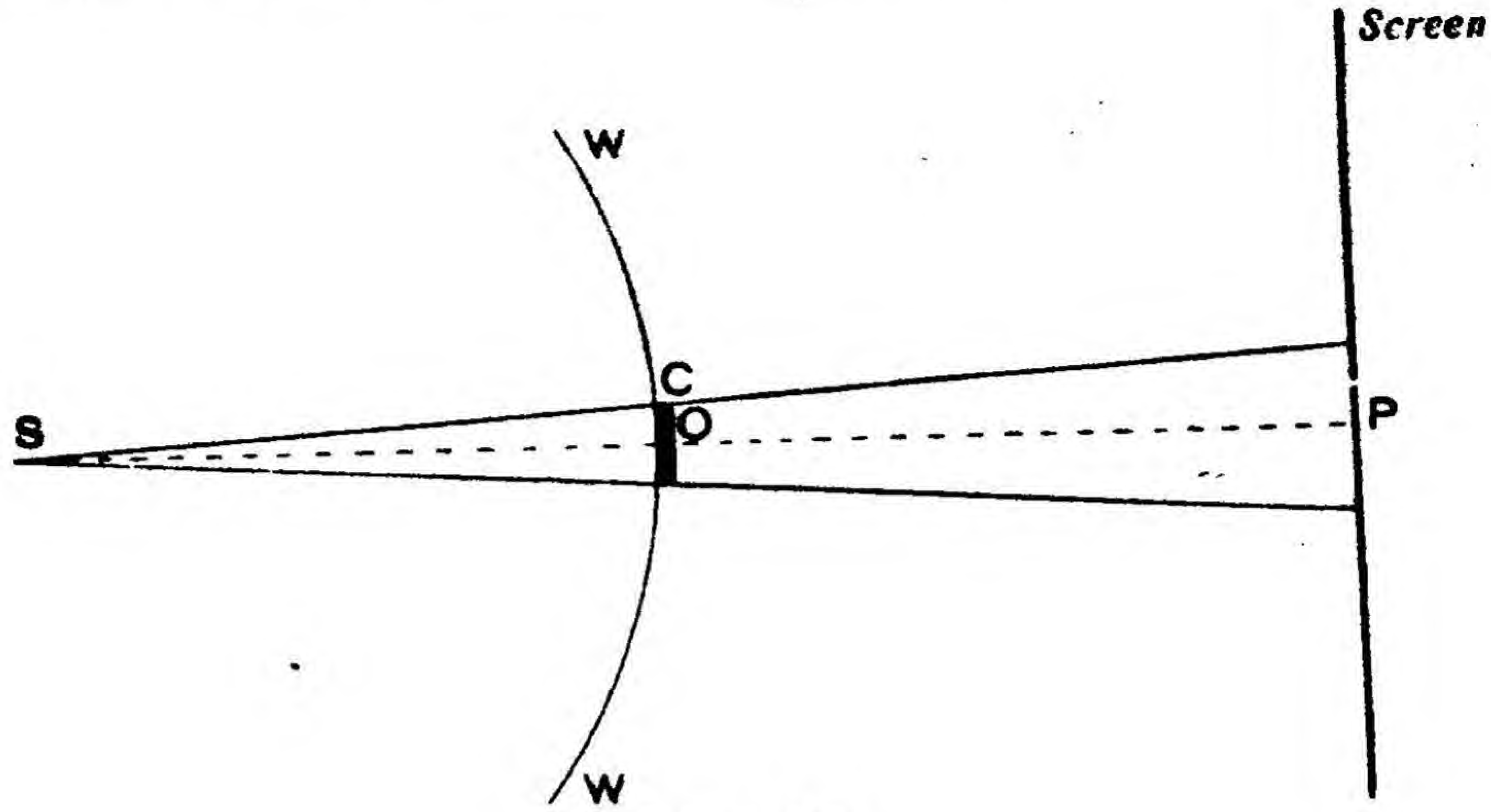


Fig. 217.

cast on a screen 2 metres away, a bright spot of light is clearly seen at the centre of it. The threepenny-piece cuts out the first 64 zones and so the spot is not very bright. It can be made brighter by moving the screen further from the obstacle. Here, then, is a clear case of diffraction, light rays bending round the corners of the threepenny-piece to come to the centre of the shadow. We can see why the bright spot comes only at the centre, because this is the one place where the secondary wavelets from the region round the edge of the obstacle are all in phase. At any other place the phase differences between the various wavelets cause the amplitude of the resultant vibration to be zero.

It is interesting to notice that Delisle had performed this experiment and noticed this result in 1715, some hundred years before it was repeated by Arago, but his result was forgotten because no one understood what it meant. It required the help of a mathematical theory of waves to interpret the result, to show that it was a case of diffraction, to show that this experiment demonstrated that the light waves did spread sideways behind an obstacle in the same way as water waves. So we see that another function of a scientific theory is to light up phenomena, as it were, to exhibit rational correlation between facts which seem superficially unrelated. Delisle's observation lay unheeded on the scrap-heap until the wave theory brought it to life after a space of a hundred years!

136. DIFFRACTION AT A CIRCULAR APERTURE

Let us now see what effects we should expect if a plane wave were passed through a circular aperture, considering in the first place the effect at a point P on the line through O, the centre of the aperture, normal to its plane (Fig. 218).. If WW represents the wave front just as it comes

to the aperture, O is also the pole of the wave front with respect to P. The aperture will let through a certain number of half-period zones. If it passes only the first zone, the amplitude at P is s_1, which is twice what it would be if the aperture were large enough to pass the whole wave front. Since the intensity of light is proportional to the square of the amplitude of the waves, *the intensity of the light at P is four times as great with this small aperture in front of the wave front as with a large aperture*! So we have the paradox that it is necessary to cover up the wave front in order to increase the brightness on the screen at P! Huxley's remark that science is nothing but trained and organised common-sense does not seem to be borne out by this statement! But the prediction

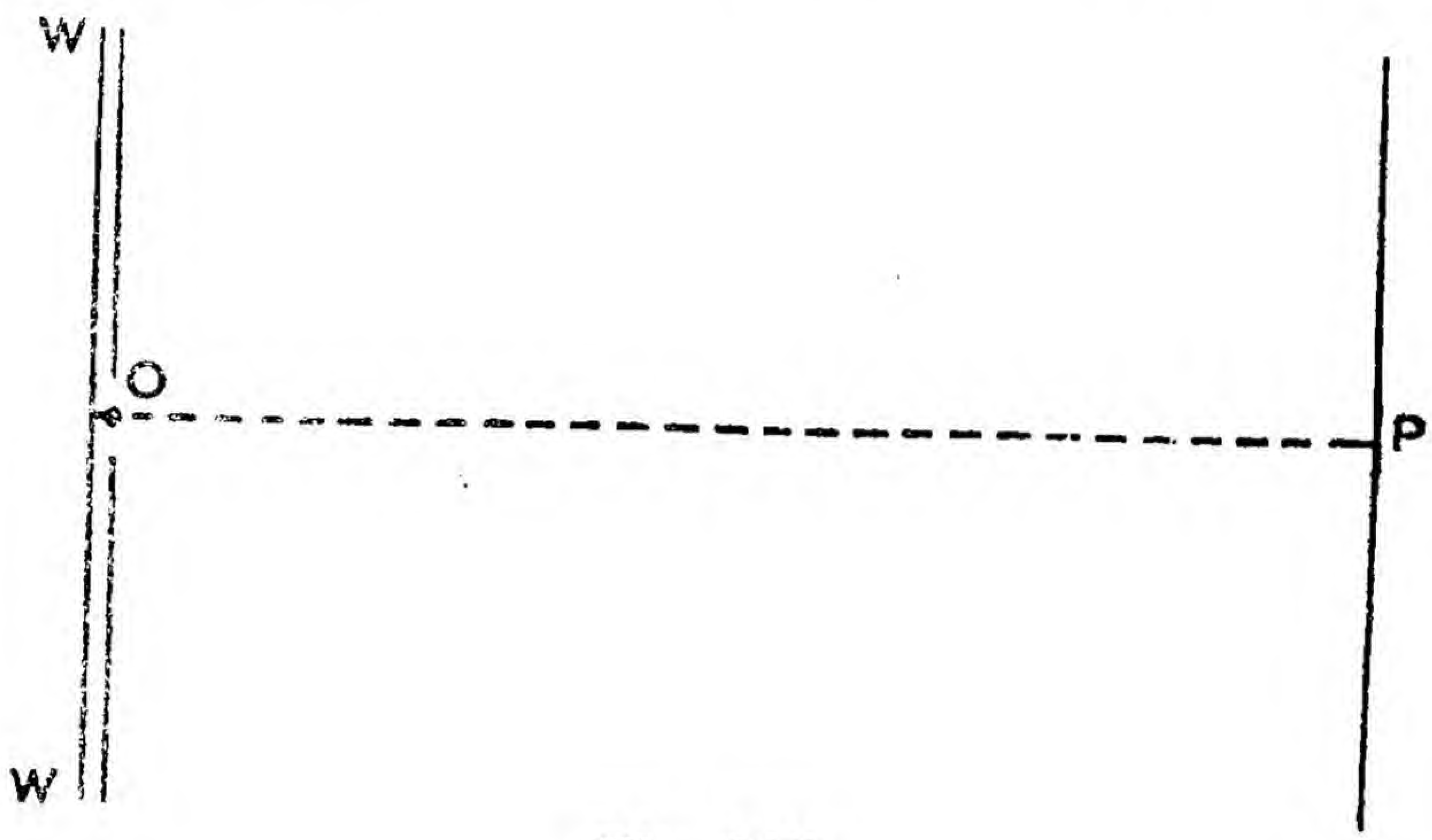


Fig. 218.

is verified by experiment. And here is a still more surprising one; if the radius of the aperture is increased so as to pass the first two zones, the resultant amplitude at P is $s_1-s_2=0$. Therefore increasing the size of the aperture causes the brightness at P to decrease to nothing! And this is verified by experiment too! In general, we can say that the illumination at P is bright or dark according as the aperture passes an odd or even number of zones.

It is not easy to make an aperture which is both strictly circular and of variable radius, but the above experiment can be varied in this way. The number of zones passed by the aperture can be changed by altering the radius of the zones, while that of the aperture remains constant. The screen is moved to be so far from the aperture that the first zone is bigger than the radius of the aperture. The illumination at P is then a maximum. The screen is then moved in towards the aperture, thus decreasing the distance *d* from the point P to the wave front and so decreasing the radii of the zones. When the screen is so near that the first two zones are passed the illumination is a minimum and there is a dark spot at P in the middle of the bright patch of light. As the screen moves still nearer a bright spot reappears at P when three zones are passed by the aperture, the dark spot having expanded to a dark circle. This process continues as the screen is moved towards the aperture, black spots

appearing at P when an even number of zones is passed by the aperture; these spots expand into circles as the screen is moved in still further. All these predictions are verified by experiment and can easily be performed in an elementary laboratory.

A strict treatment of the illumination at points off the line OP requires the mathematical method of adding up the amplitudes of the wavelets and is beyond the scope of this book, but we can see the reason for the dark circles in the following way. When the screen is near enough to the aperture for it to pass three zones, P is bright. But, as we move up along the screen from P, the pole of the point whose illumination we are considering moves up from O towards the edge of the aperture. When the pole reaches the point such that only the first two zones are passed completely, then part of the third and part of the fourth zones will also be passed. The first two zones cancel each other out and there is a point at which the parts of the third and fourth zones passing the aperture cancel out too. When this is so, the effect at the corresponding point on the screen above P is a minimum. Since the apparatus is symmetrical about the line OP as axis, this effect will produce a dark ring with P as centre.

137. DIFFRACTION AT A STRAIGHT EDGE

Both the cases of diffraction which we have considered have had axial symmetry, for the object or aperture has been circular and the waves have come ultimately from a point source. We now turn to cases in which the source is a slit, ideally of infinite length, emitting cylindrical waves with the slit as axis and the diffracting objects are edges, narrow obstacles, or slits parallel to the slit source.

We must first see how to add together the secondary wavelets from a cylindrical wave front when they reach a given point. Let S represent a slit source normal to the plane of the paper and let WW represent a section of the cylindrical wave front emitted by the source (Fig. 219). We shall find the effect of the whole wave front at a point such as P by dividing it up into zones the amplitudes of whose wavelets at P are alternately positive and negative. Join SP to cut the wave front in O and divide the wave front into two equal halves by a line through O normal to the plane of the paper. With centre P and radii $d+\frac{\lambda}{2}$, $d+\frac{2\lambda}{2}$, $d+\frac{3\lambda}{2}$, $d+\frac{4\lambda}{2}$ draw a set of circles cutting WW in the points AA, BB, CC, DD, and so on. If lines are drawn through these points normal to the plane of the paper, they divide each half of the wave front into a set of **half-period strips**. Consider the first half-period strip AO in the upper half of the wave front. The wavelets from it reach the point P later as we go further from the line AO up or down the strip in a direction

normal to the plane of the paper. Each half of the strip must be subdivided along its length into a set of half-period elements by drawing circles centre P radii $d+\frac{\lambda}{2}$, $d+\frac{2\lambda}{2}$, $d+\frac{3\lambda}{2}$, and so on, in the plane through OP normal to the plane of the paper. The wavelets from the first half-period element will be out of phase with those from the second element

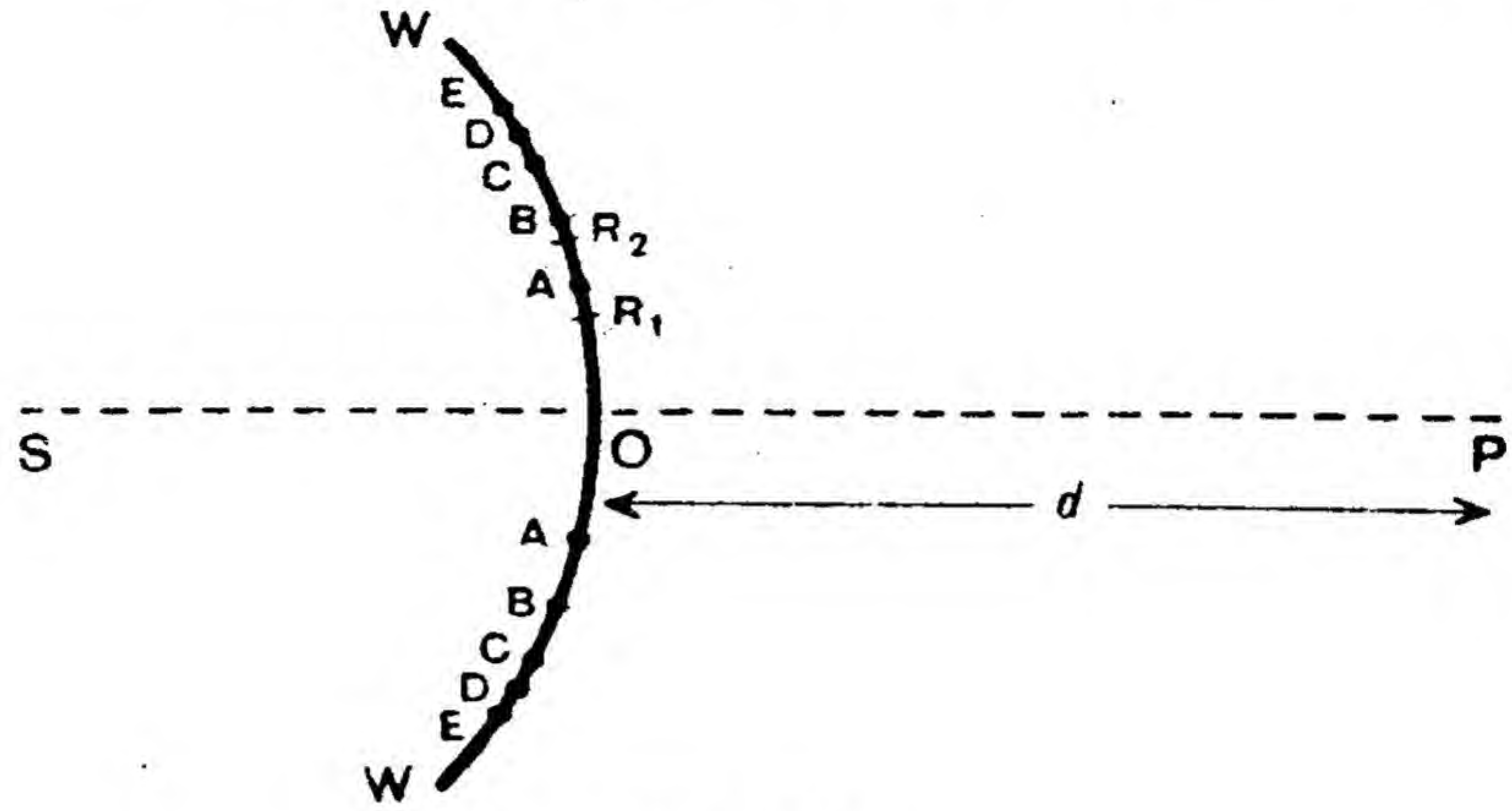


Fig. 219.

at the point P, since the distance of corresponding points in the two elements from P differs by half a wave-length. So the total effect of each half of the first half period strip can be found as with a plane wave and works out to be half the effect of the first half-period element. The same is true of all the other half-period strips. In other words, we can ignore all the wave front above and below the plane of the paper except the region in the immediate neighbourhood of the line WW. Now we have deliberately constructed the half-period strips so that a crest from any point R_1 in the first half-period strip reaches P at the same instant as a trough from a corresponding point R_2 in the second strip. The whole of the two strips can be dealt with in this way and, if the amplitude of the wavelets from the first half-period strip at P is called s_1, that from the second, third, fourth strip and so on will be $-s_2$, $+s_3$, $-s_4$, and so on. A more exact analysis shows that the numerical values of these amplitudes fall off more rapidly than in the case of a plane wave front. The reader must remember that there are two halves of the wave front to be taken into account now. It follows in the same way as with the plane wave front that, if there is no obstacle, the amplitude of each half of the wave front at P is $\frac{s_1}{2}$ and so the amplitude of the wavelets from the whole wave front is s_1.

Let us now consider the effects to be expected if a slit source through S normal to the plane of the paper (Fig. 220) sends a cylindrical wave front on to a straight edge through E normal to the plane of the paper, the edge of the geometrical shadow on the screen on which the light falls passing through G on the line SE produced. The illumination at a point P_1 outside the geometrical shadow is found by dividing the wave front into

two halves by the line SP cutting it at O_1. The amplitude of the light at P_1 is that due to the upper half of the wave front, which is entirely unaffected by the straight edge, together with that from the portion O_1E of the wave front. The amplitude at P_1 will therefore be a maximum or minimum according as O_1E contains an odd or even number of half-period strips, that is, according as $EP_1 - O_1P_1$ is equal to $(2n+1)\frac{\lambda}{2}$ or $n\lambda$, where n is an integer. As P_1 moves up the screen from G, the illumination will pass through a set of maxima and minima as the number of half-period strips in O_1E becomes 1, 2, 3, 4, and so on. So we shall see a

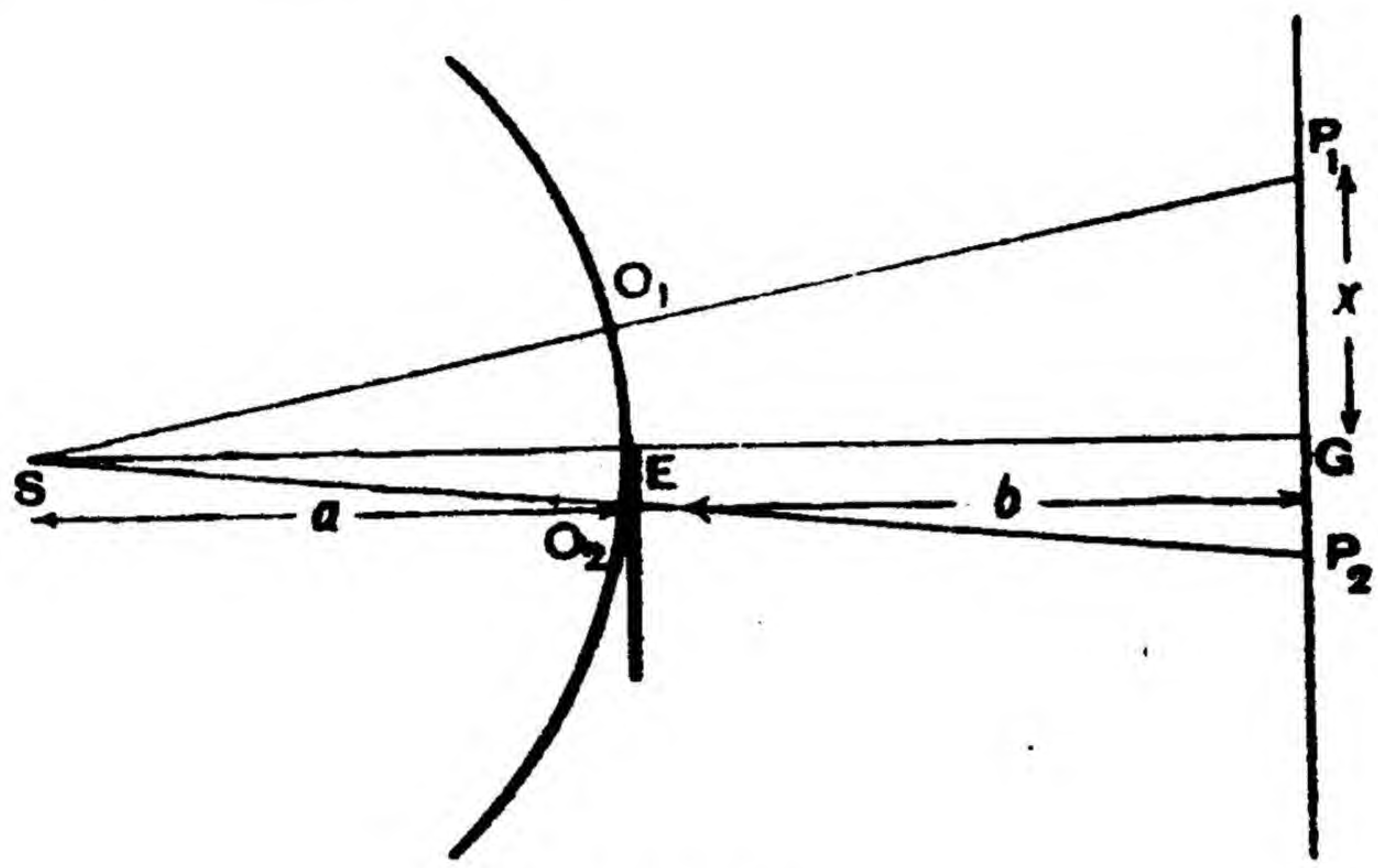


Fig. 220.

set of alternate bright and dark bands parallel to the edge of the geometrical shadow, but the dark bands will not be completely dark, since the upper half of the wave front always sends light to this part of the screen.

What can we deduce about the width of these diffraction bands? If P_1 is at the n^{th} dark band

$$EP_1 - O_1P_1 = n\lambda.$$

$$EP_1 = (b^2 + x^2)^{\frac{1}{2}}$$

$$= b\left(1 + \frac{x^2}{b^2}\right)^{\frac{1}{2}}$$

$$= b\left(1 + \frac{1}{2}\frac{x^2}{b^2}\right)$$

expanding by the binomial theorem and neglecting the higher terms.

$$\therefore\ EP_1 = b + \frac{1}{2}\frac{x^2}{b}$$

Similarly
$$SP_1=(a+b)+\frac{x^2}{2(a+b)}$$
$$\therefore O_1P_1=b+\frac{x^2}{2(a+b)}$$
$$\therefore EP_1-O_1P_1=\frac{x^2}{2b}-\frac{x^2}{2(a+b)}$$
$$=\frac{x^2a}{2b(a+b)}$$
$$\therefore \frac{x^2a}{2b(a+b)}=n\lambda$$
$$\therefore x=\sqrt{n\,.\,\frac{2b(a+b)}{a}\,.\,\lambda} \quad . \quad . \quad . \quad . \quad . \quad . \quad (76)$$

and so the distances of the dark bands from the edge of the geometrical shadow are proportional to the square roots of the natural numbers and they get closer together as we go out from the shadow. This fact, together with their poor contrast owing to the minima not being absolute, enables them to be distinguished from interference fringes.

The illumination at a point P_2 inside the geometrical shadow is due entirely to wavelets from the upper half of the wave front, since the middle of the wave front O_2 is behind the edge E. If the obstacle cuts out the first r strips, the effect at P is $\frac{S_{r+1}}{2}$. This rapidly diminishes to zero as r increases, so that the illumination below G rapidly falls off to zero. The graph of intensity of illumination on the screen against distance from the

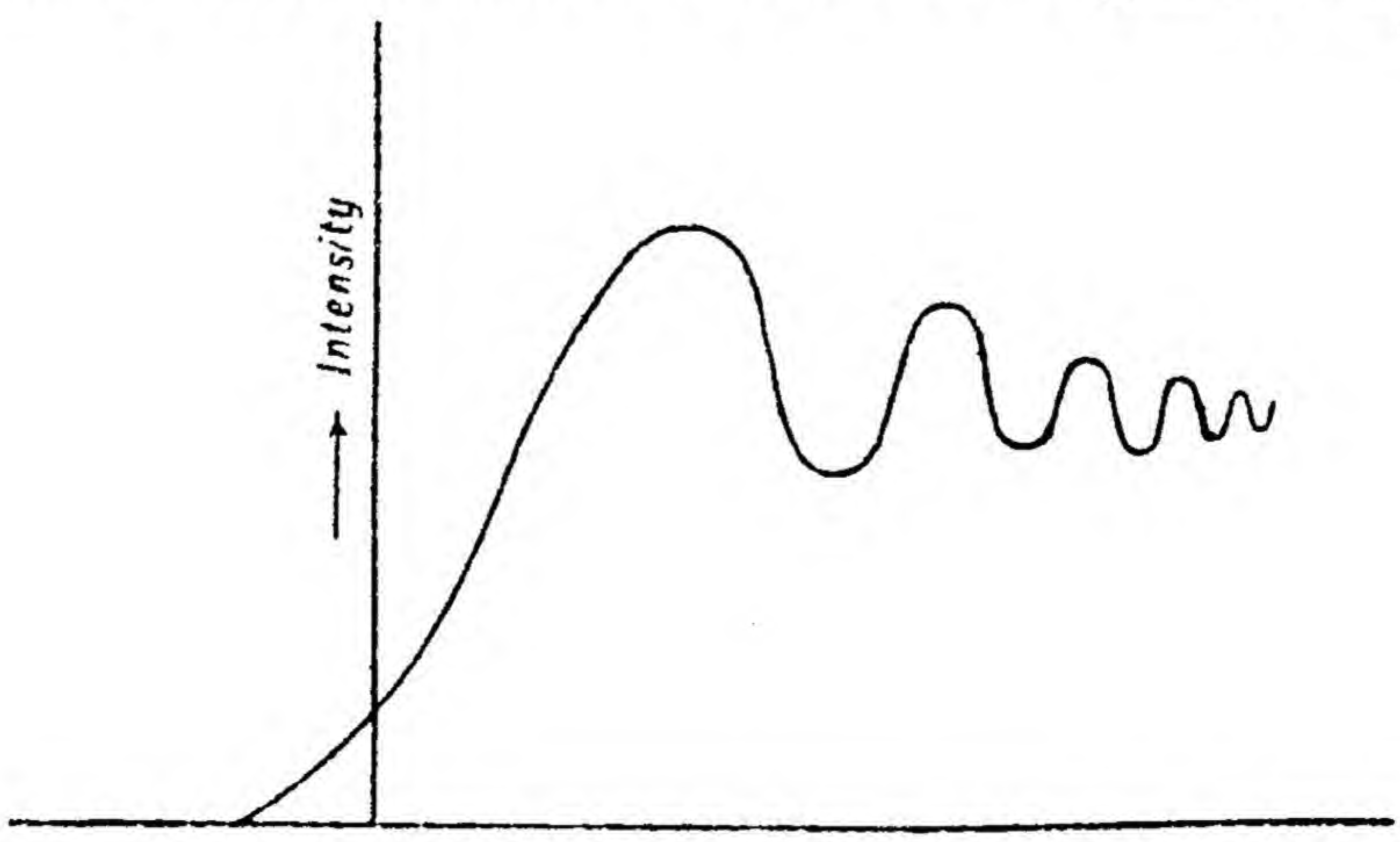


Fig. 221.

edge of the geometrical shadow is shown in Fig. 221, from which we see that the edge of the shadow is not sharp as it would be if rectilinear propagation were strictly true, and also the light does spread sideways a little behind the corner of the obstacle, as the region of complete darkness begins at a finite distance below G. All these effects are observed

in practice using a straight edge about 2 metres from an illuminated slit, the screen being some 2 metres on the other side of the edge from the source. Finally, if the slit is illuminated with sodium light and the fringes outside the shadow are measured, the wave-length of light can be determined and the value agrees with that found by the various interference methods. *It is this quantitative prediction of the unequal spacing of the fringes, its verification, and the measurement of the wave-length of light, which is such convincing evidence that we are indeed dealing with diffraction, although the actual amount of bending or sideways spreading of the light is so small.* It is not too much to say that the supporters of the corpuscular theory would never have accepted these phenomena as evidence of the diffraction of light waves by mere analogy with what happens with water waves, since the effects are so much smaller and so different in character. It requires a theory of waves to show that these apparently insignificant effects close to the edge of shadows are the key to the true nature of light.

138. DIFFRACTION AT A NARROW OBSTACLE

We are now in a position to discuss the classical case of the shadow of a narrow obstacle observed by Grimaldi and Newton with such care, but unexplained by them. S represents an illuminated slit (Fig. 222) sending light on to a narrow obstacle represented by AB, the shadow being cast on a screen some distance beyond the obstacle. The obstacle can be a copper wire about 1 mm. in diameter and GG represents the geometrical shadow. The effects at a point such as P_1 outside the geometrical shadow will be just the same as those at a straight edge, since

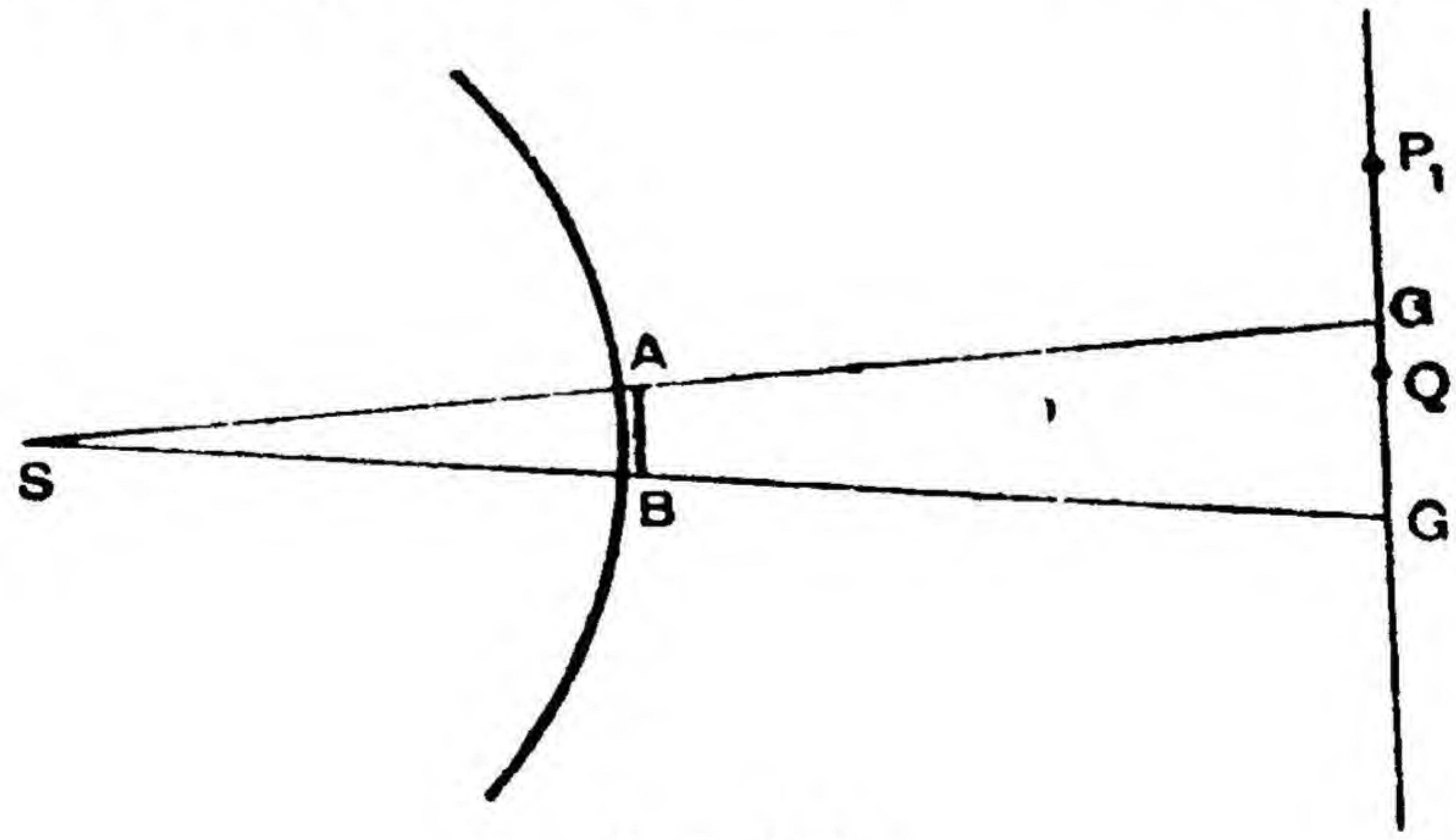


Fig. 222.

the wavelets reaching P_1 from the lower half of the wave front will have a negligible effect owing to the finite breadth of the obstacle. So we shall get unequally spaced dark and bright fringes parallel to the edge of the obstacle one each side of the shadow. The illumination at a point such as Q inside the shadow is simply half the effect of the first half period strip on either side of the obstacle. These strips are so narrow that they can be regarded as a narrow slit on each side of the obstacle

and the effect at Q will be a maximum or minimum according as BQ—AQ is $n\lambda$ or $(2n+1)\frac{\lambda}{2}$. In fact, the fringes will be just the same as with Young's Slits and will be equally spaced of width $\frac{d\lambda}{s}$, where s is now the width of the obstacle AB and d is its distance from the screen. The reader will recall (Art. 97) that this is precisely what Grimaldi saw, so his observations are strikingly verified by the wave theory. (Plate VI.) Possibly Newton failed to get any fringes inside the shadow because his obstacle was too wide, in which case the fringes would have been too close together to be seen with the naked eye. Yet again we see how theory brings observations to life by showing their significance in a rational scheme. This also shows the uselessness of isolated observations. Neither Newton nor Grimaldi, both of whom were familiar with diffraction in sound and water waves, saw that the above effects were diffraction. It is not surprising that they did not, as the effects were so much feebler and so different in character than those obtained with sound and water waves, where the ordinary conditions were so different from the above conditions in light.

139. DIFFRACTION AT A NARROW APERTURE

It is natural to pass on to the case of a narrow aperture AB (Fig. 223) illuminated by light from a slit S normal to the plane of the paper, the effects being cast on a screen some distance beyond the aperture. GG marks the edges of the patch of light to be expected on rectilinear propagation and we shall consider the illumination at its centre O, at a point such as P at a distance from the centre but still in the patch of light on the screen, and finally at a point Q inside the geometrical shadow. The illumination at O is a maximum or minimum according as the aperture passes an odd or even number of half-period strips from *each half* of the wave front. As the aperture is increased in width from a very small value, or as the screen is moved up to it from a very large distance, the illumination at O starts by being very large, since only one or a portion of one half-period zone passes through the aperture from each half of the wave front. Then it goes through a set of minima and maxima, until it gradually becomes uniform when a large number of half-period strips in each half of the wave front passes through the aperture. The total amplitude is then $\frac{s_1}{2}+\frac{s_1}{2}=s_1$, where s_1 is the amplitude at O of the wavelets from the first half-period strip of one half of the wave front.

Let us now suppose that the aperture is so wide as to pass, say, a hundred half-period strips from each half of the wave front and we wish to find the illumination at a point such as P. Let us start with P at O and let it move up towards G along the screen. A point will be reached

when 99 half-period strips from the upper half and 101 from the lower half of the wave front are passed by the aperture, but this will make no difference to the illumination at P, since the amplitude of each half at P will still be $\frac{s_1}{2}$. This will continue until we get so close to G that only ten or so half-period strips from the upper half of the wave front pass through the aperture, when we shall evidently get maxima or minima at P according as an odd or even number of half-period zones get through the aperture. In fact, we shall get precisely the same fringes as with a straight edge, and, as in that case, the illumination quickly falls off to zero inside the geometrical shadow. The same thing applies to the other part of the geometrical shadow and so a wide aperture behaves like two independent straight edges.

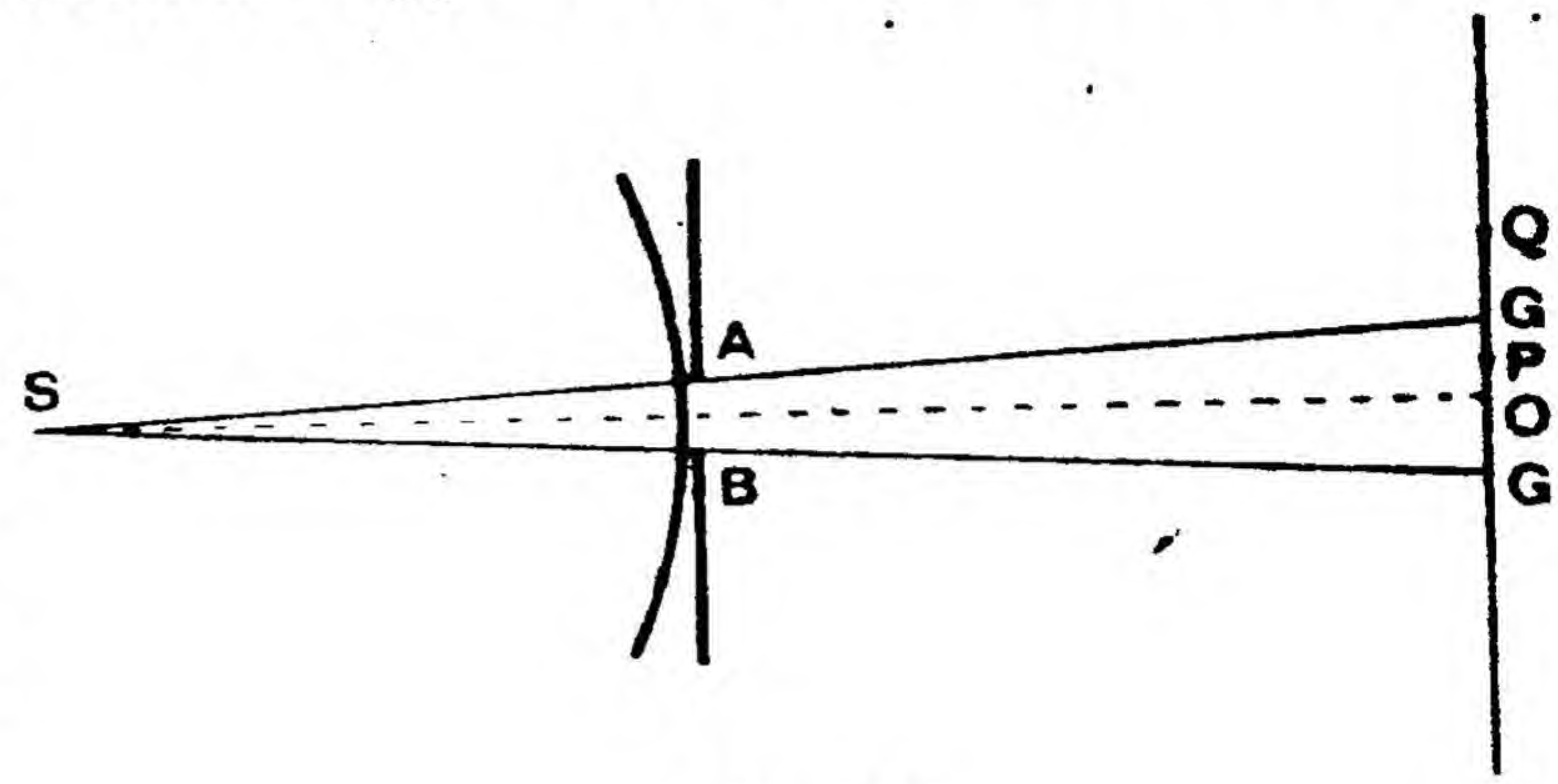


Fig. 223.

Let us now suppose that the aperture is so narrow that only five half-period strips from each half of the wave front pass through it, the illumination at O being a maximum. As P moves up from O, a point is reached when four strips from the upper half and six from the lower half pass through the aperture. The amplitude of the wavelets from the four strips is $s_1+s_2-s_3+s_4=0$, while that from the six strips is $s_1-s_2+s_3-s_4+s_5-s_6=0$. So the illumination at P is a minimum. When P is a little nearer to G, three strips from the upper half and seven strips from the lower half will be able to send wavelets through the aperture. The amplitude of the three strips at P is $s_1-s_2+s_3=s_3$, and that of the seven is s_7, and so the illumination at P is now a maximum. In general as P moves towards O, the illumination is a maximum or minimum according as an odd or even number of half-period strips from each half of the wave front sends wavelets to the point P. So there will be a set of alternate dark and bright bands parallel to the edges of the geometrical shadow inside the patch of light on the screen. Finally consider a point such as Q inside the geometrical shadow. The illumination here is entirely due to one half of the wave front and will be a maximum or minimum according as the aperture includes an odd or even number of half-period strips, that is, according as QB−QA is $(2n+1)\frac{\lambda}{2}$ or $n\lambda$. This is a similar condition to that governing the maxima and minima in Young's Slits only here we get a *maximum* when the path difference from the edges of the aperture to the point is an *odd* number of half wave-lengths instead

of an even number as in Young's Slits. The reason for the difference is that in Young's Slits we have two point or line sources and only two interfering wave trains, whereas here we have a section of wave front of finite width and an infinite number of interfering wave trains. We can see the truth of the above result in another way by dividing the wave front passing through the aperture into two halves. If QB$-$QA is λ, for example, then the path difference for the wavelet from B and the one from the middle of the aperture is $\frac{\lambda}{2}$, and so they cancel each other out at Q. Similarly a wavelet from any other point in the lower half of the wave front is cancelled out by one from the corresponding point in the upper half of the wave front, the point being such that it is $\frac{\lambda}{2}$ nearer to Q. Hence the wavelets from the lower half of the wave front are cancelled out by those from the upper half of the wave front and the illumination at Q is a minimum. The condition that we get a maximum or minimum at Q according as QB$-$QA is $(2n+1)\frac{\lambda}{2}$ or $n\lambda$ means that we get alternate dark and bright bands parallel to the edge of the geometrical shadow, equally spaced and of width $\frac{d\lambda}{s}$, where d is the distance of the screen from the aperture and s is the width of the aperture. All these effects are verified by experiment. So we may now regard the diffraction of light as established beyond doubt. It is time to turn to some uses which have been made of this diffraction and we start with a rather delightful example of it, which is pleasing rather than useful. It shows how diffraction can be made to produce the same effect as a lens.

140. THE ZONE PLATE

Let A be a point source of light emitting spherical waves, whose effect at the point B is to be found (Fig. 224). Draw a plane through O normal to the plane of the paper and divide it into zones bounded by circles centre O and radii OP_1, OP_2, OP_3, and so on, where $AP_1+BP_1=AO+BO+\frac{\lambda}{2}$, $AP_2+BP_2=AO+BO+\frac{2\lambda}{2}$, and so on. These are half-period zones exactly similar to those constructed for a plane wave front (Art. 134) and crests of the wavelets from the first zone will reach B at the same time as troughs of the wavelets from the second zone and so on. Therefore, if the amplitude of the wavelets from the first zone at B is called s_1, that from the second, third, and succeeding zones will be $-s_2$, s_3, $-s_4$ and so on. If the even zones are rendered opaque to light, the amplitude at B of all the secondary wavelets starting from the plane at O is

$s_1+s_3+s_5+s_7+s_9 \ldots$, which is a maximum. In fact, B is a kind of image of A. But an interesting consequence follows from a calculation of the radii of the zones. The radius, r_n, of the n^{th} zone is given by

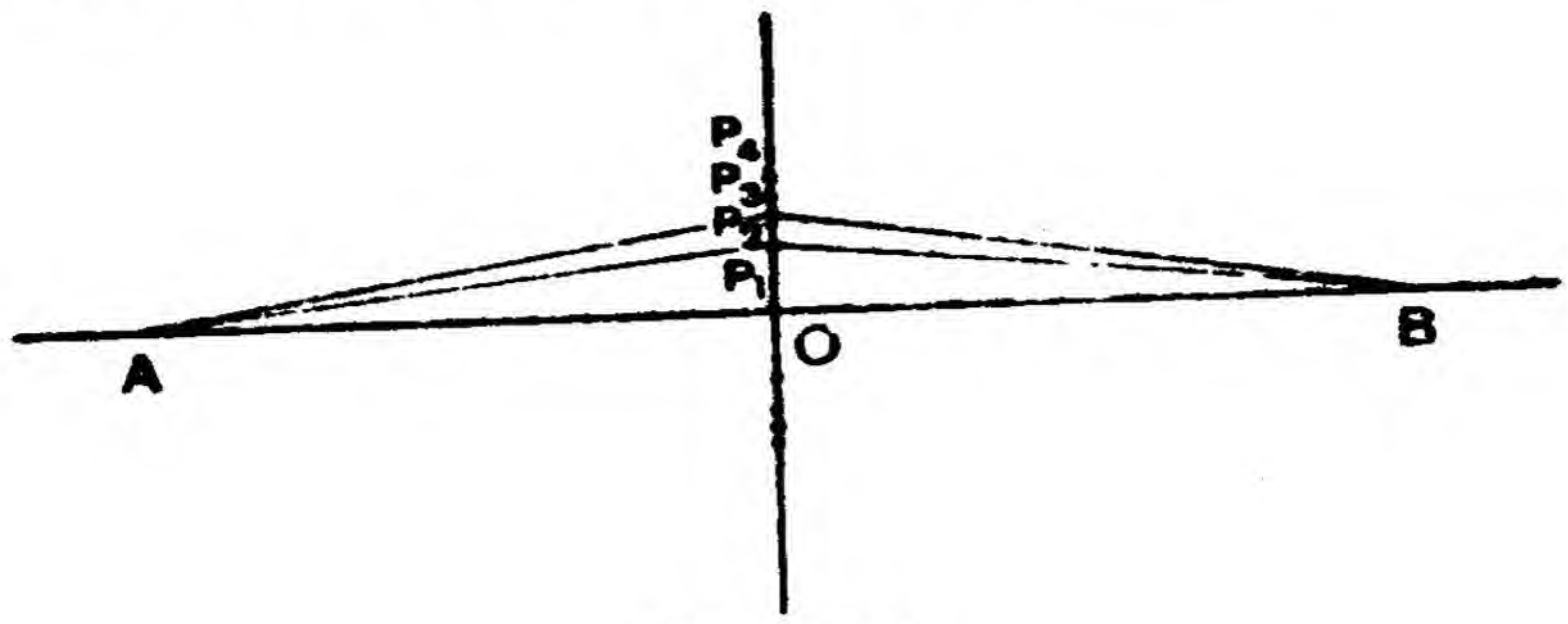


Fig. 224.

$$(AO^2+r_n^2)^{\frac{1}{2}}+(BO^2+r_n^2)^{\frac{1}{2}}=AO+BO+\frac{n\lambda}{2}$$

$$\therefore \; AO\left(1+\frac{r_n^2}{AO^2}\right)^{\frac{1}{2}}+BO\left(1+\frac{r_n^2}{BO^2}\right)^{\frac{1}{2}}=AO+BO+\frac{n\lambda}{2}$$

Expanding by the binomial theorem and neglecting all but the first two terms,

$$AO+\frac{r_n^2}{2AO}+BO+\frac{r_n^2}{2BO}=AO+BO+\frac{n\lambda}{2}$$

$$\therefore \; r_n^2\left(\frac{1}{AO}+\frac{1}{BO}\right)=n\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (77)$$

so that the radii of the circles are proportional to the square roots of the natural numbers, as was the case with Fresnel's zones for a plane wave front. But we can re-arrange this equation in the form

$$\frac{1}{AO}+\frac{1}{BO}=\frac{n\lambda}{r_n^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (78)$$

which is identical with the equation relating the *distance* of a real object with that of its real image formed by a converging lens. In fact, this device behaves like a converging lens, the numerical value of whose focal length is $\frac{r_n^2}{n\lambda}$.

Such a device is called a zone plate, and the way in which it is made will be understood when it is realised that the radius of the first zone in a zone plate of focal length 20 cm. is, from equation (78), $\sqrt{20\times5\times10^{-5}}$, or 0·316 mm., taking λ as 5×10^{-5} cm. It is evidently impossible to rule the zones directly on a glass plate, so a set of circles whose radii are in the ratio of the square roots of the natural numbers is carefully drawn on white paper, the radius of the first zone being some 3 cm. The odd zones are then painted black and an image of these alternate black and white zones diminished about 100 times is cast on a photographic plate

and, when it has been fixed, the negative is the zone plate with the odd zones transparent and the even zones opaque to light. If the radius of the first zone is 0·316 mm., the plate acts like a lens of focal length 20 cm. and will form images in the same way as that lens. It is interesting to observe that Rayleigh suggested that the intensity of the image could be increased fourfold if the even opaque zones were replaced by zones of such thickness that the light was retarded just half a wave-length in passing through the plate. This has actually been accomplished by an American physicist, R. W. Wood, whose book on Physical Optics should be consulted for details of the process. Again, the zone plate has a second focal length produced when wavelets from the 3rd, 7th, 11th, . . . zones add together, and this focal length is three times as great as that produced when all the zones add together. All these effects have been realised experimentally and form a striking confirmation of the existence of diffraction and Fresnel's theory to account for it.

141. THE FRAUNHOFER CASES OF DIFFRACTION

The cases of diffraction which we have considered so far are in the Fresnel class, in which the source of light is at a finite distance from the system producing the diffraction and effects are sought at a definite *point* on a screen at a finite distance from the diffracting system. We shall now discuss the other important class of diffraction effects, the Fraunhofer class, in which the source is at infinity and effects are sought in a definite *direction*. Since the wavelets producing the effect in a given direction are parallel to one another, they are made to cross by sending them through a lens, when they cross at a point in its focal plane, where the effect is to be seen. This type of diffraction is best produced by putting the diffracting system on a spectrometer table, the collimator producing the parallel light which falls on the system and the telescope objective bringing the wavelets diffracted in a given direction to a focus at the same point. Hence the effects are seen by looking down the spectrometer telescope in the usual way.

We shall consider first of all the case of diffraction at a single slit. AB represents a horizontal cross-section of a vertical slit mounted on the table of a spectrometer, parallel light of wave-length λ falling on it from the collimator (Fig. 225). We know from our experience of water waves that most of the light goes straight on, as shown by the dotted lines, but some of it will bend round the corners of the slit and spread out in all directions. We will consider the rays diffracted through an angle θ, that is, those rays which leave the slit at an angle θ to those which go straight on. In order to make these rays cross at the focus of the objective L of the spectrometer telescope, we turn the telescope round so that its axis makes an angle θ with the normal to the plane of the slit as shown in Fig. 225. A given crest in the incident waves reaches A and B simultane-

ously and, if BD is drawn normal to the rays diffracted through an angle θ, the time taken by rays to go from B to F is the same as that taken to go from D to F. So the path difference between crests starting from A and B at the same time and crossing at F, where the interference can

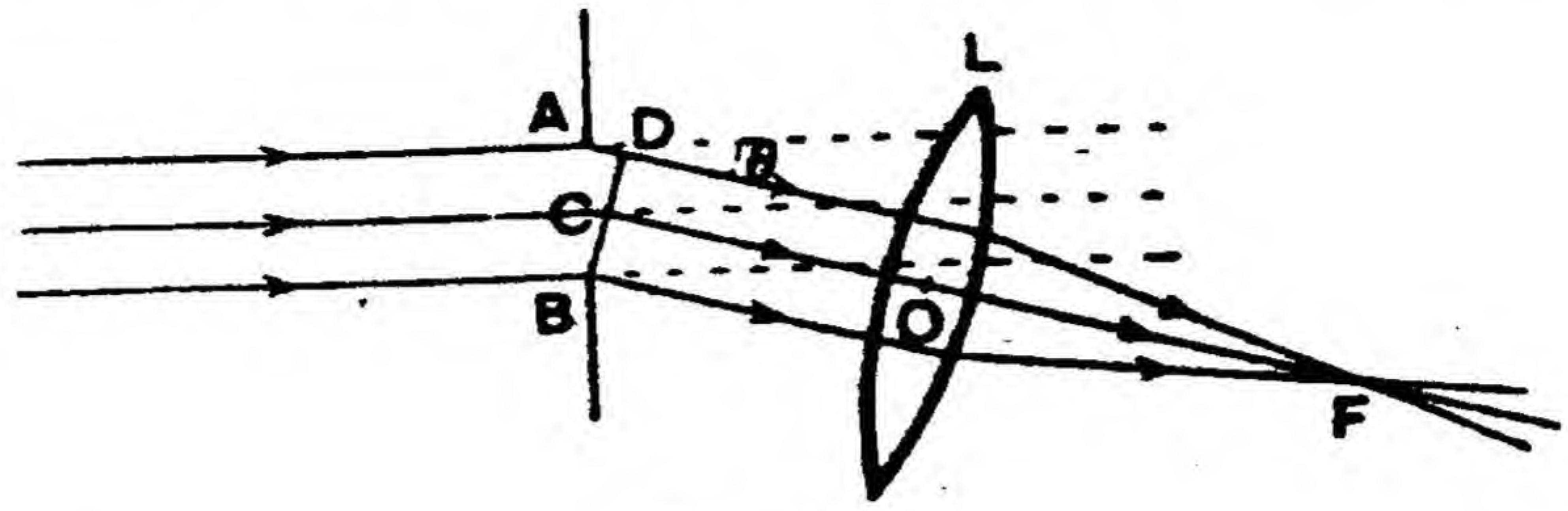


Fig. 225.

be seen, is AD, which is $a \sin \theta$, where a=AB, the width of the slit. If this path difference is λ, there will be a ***minimum*** at F, since the path difference between the rays from two corresponding points such as A and C, half the width of the slit apart, is $\frac{\lambda}{2}$ and the rays annul one another at F. Hence the slit can be divided into two equal halves, such that the rays from one half annul the corresponding rays from the other half at F; therefore the effect of the whole slit at that point is zero. Minima also occur if the path difference is $n\lambda$, where n is an integer, as can be shown by dividing the slit into two halves if n is an odd number, or into a sufficient number of portions to give a path difference equal to an odd number of half wave-lengths between corresponding points if n is an even number. Maxima, whose intensity is considerably less than the main maximum, occur between these minima. So we may sum up the effect to be expected in this way. The **diffraction pattern** of the slit is a central maximum in the direction of the incident light and a set of alternate maxima and

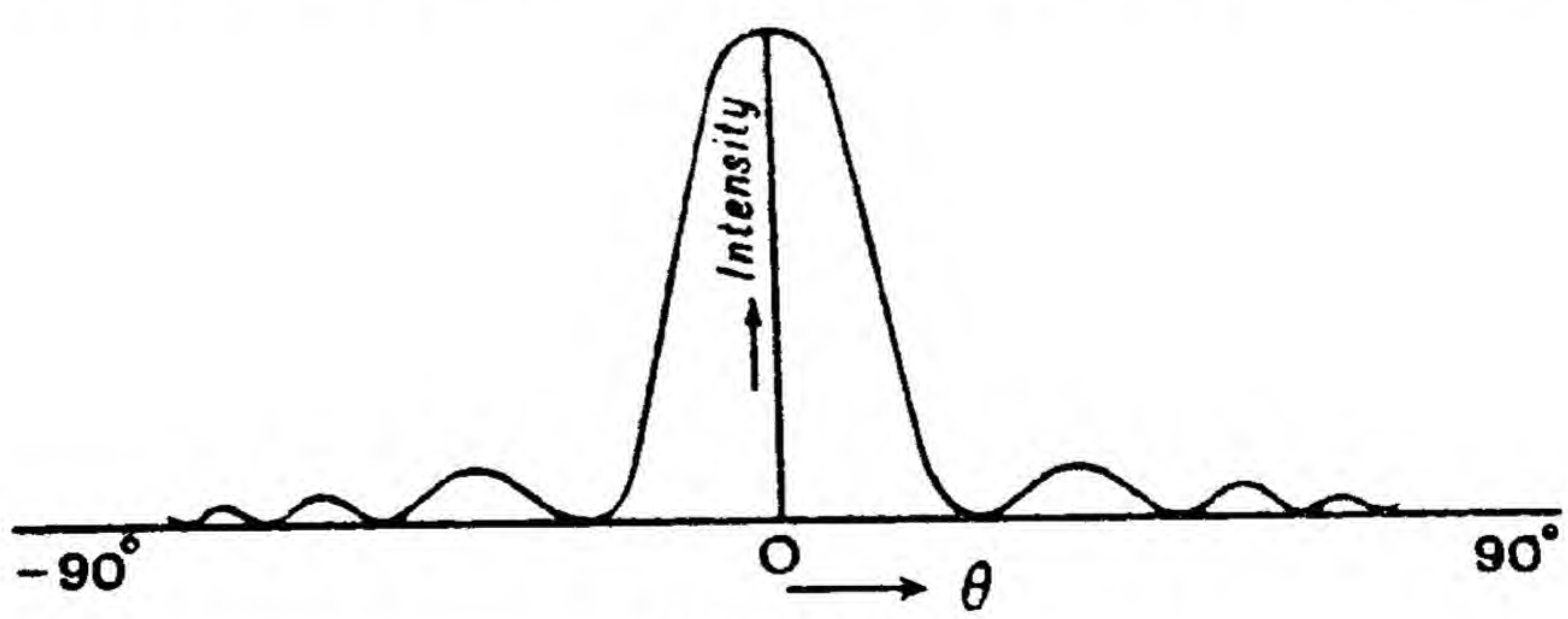


Fig. 226.

minima on either side of the central maximum, the minima being in directions given by the equation

$$a \sin \theta = n\lambda$$

The statement that a minimum lies in a given direction means that it is at the focus of the telescope objective, when its axis is parallel to that direction. Each maximum is a slit parallel to the actual slit. So we get,

as it were, a main image of the slit with a set of fainter subsidiary images on either side of it, the relation between intensity of illumination and direction being shown in Fig. 226. These predictions are completely verified by experiment.

142. THE DIFFRACTION GRATING

We now consider the diffraction effects produced by a set of parallel equidistant slits. Such an arrangement is called a diffraction grating and it can be used to measure the wave-length of light very accurately. The first gratings were made by Fraunhofer and consisted of silver wires stretched on a frame, there being about 200 wires to the centimetre. Then gratings were made by ruling lines with a diamond on a glass plate, the slits being the untouched parts of the glass between the lines, which were effectively opaque to light. Finally gratings were ruled on speculum metal with as many as 5,000 lines per centimetre. These gratings are difficult to make and are therefore very expensive, but celluloid casts or replicas of them can be manufactured for three pounds and are used in many laboratories.

The effects produced by a diffraction grating are in the Fraunhofer class and so it is mounted on the table of a spectrometer which has been adjusted in the usual way. A horizontal cross-section of the grating is shown in Fig. 227, AB, CD, GH, JK, . . . representing slits normal to the

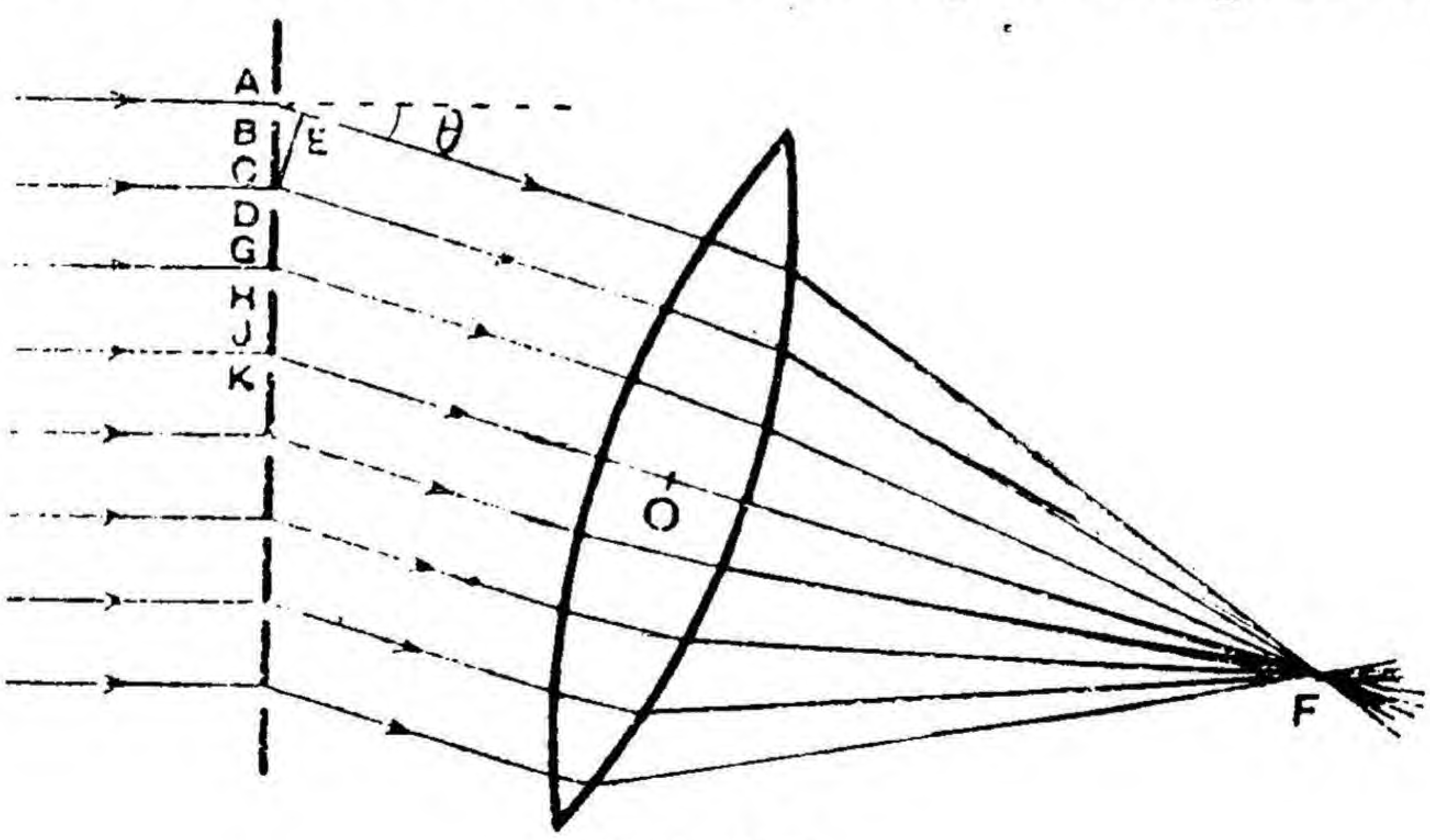


Fig. 227.

plane of the paper, while BC, DG, HJ . . . represent the opaque lines which have been produced by the ruling point. Let a parallel beam of monochromatic light of wave-length λ fall normally on the grating. Most of the light issuing from the slits will go straight on, but, as the width of each slit is of the order of the wave-length of light, some of the light will spread out on leaving each slit. Let us consider the rays diffracted through an angle θ, which fall on the telescope objective and will cross at its focus F, when it has been turned so that its axis makes an angle θ with the light which has gone directly through the grating. If

we draw the rays from the edges A, C, G, J, . . . of successive slits and draw CE at right angles to the ray AE from A, the path difference between the rays from A and C, which will cross at F, is AE, since the rays are in phase at A and C and the light takes the same time to go from C to F as from E to F. But AE$=e \sin \theta$, where $e=$AC and is called the **grating element**. Therefore there will be a maximum at F for those values of θ given by the equation

$$e \sin \theta = n\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (79)$$

where n is an integer, since the rays from A, C, E, J, . . . will reinforce one another at F, and so will those from the other corresponding points of the slits. Therefore, as the telescope of the spectrometer is turned from the position in which its axis is normal to the grating, the focus of the objective comes to the first place of maximum illumination at the value of θ given by $e \sin \theta = \lambda$, to the second place at the value of θ given by $e \sin \theta = 2\lambda$, and so on. At each of these positions an image of the slit falls on the vertical cross-wire of the telescope. Similar images are formed on the other side of the direct beam and they are called the first, second . . . order diffracted images respectively.

If white light is substituted for monochromatic light, it follows from equation (79) that a spectrum will be formed in the focal plane of the objective of the telescope, since there is a different value of θ for each different colour or wave-length for a given value of n. The reader will realise that, for one position of the telescope, only one colour comes to a focus at the actual focus of the objective. The other colours send parallel beams making different angles with the axis of the telescope, and, corresponding to each direction, there is a definite point focus in the focal plane of the objective. In this spectrum the violet, having a smaller wave-length than the red, is deviated less than the red, the opposite being the case in the spectrum produced by a prism.

The wave-length of, say, sodium yellow light is measured in the following way. The spectrometer is adjusted in the usual way, so that the collimator produces parallel light and the telescope brings it to a focus at the cross-wires, on which the eyepiece is focussed. The grating is then mounted on the spectrometer table so as to be normal to the incident light, the slit of the collimator is illuminated with sodium light from a sodium flame or a sodium lamp, and the vertical cross-wire of the telescope is set on the image of the slit produced by light which has come straight through the grating. The reading of the telescope on its scale is noted and it is then turned until the vertical cross-wire coincides with the first diffracted image of the slit. This will look yellow and, if it is not sharp, its sharpness can be improved by rotating the grating in its own plane to bring the slits of the grating parallel to the slit of the collimator. When this has been done it will be seen that there are two images of the slit very close together if a grating with about 5000 lines to the centimetre and some 2 or 3 cm. wide is used. This shows that sodium yellow light

is not strictly monochromatic, but consists of two wave-lengths very nearly equal. They are called the D_1 and D_2 lines and the wave-length of each is measured in this experiment. The vertical cross-wire is set on the centre of each line in turn, the position of the telescope being read in each case. The difference between this position of the telescope and its direct position gives the angle of diffraction for each of the first order images. The telescope is then set on the corresponding images on the other side of the direct beam and another value of each angle of diffraction is obtained, the mean of these two values being taken as the true angle of diffraction for each line. The reading on each side of the direct beam is taken in case the grating is not quite normal to the incident light. The wave-length of each line is then calculated from the equation

$$e \sin \theta = \lambda$$

e being obtained from the number of lines per centimetre of the grating, which is usually supplied by the makers. Further values of the wave-length are obtained by measuring the angles of diffraction of the second order images and also the third order, if they are bright enough for accurate measurement, and the mean of the values of the wave-length of each line found in this way is taken as its true wave-length.

The sort of angles of diffraction obtained can be seen if we assume that the grating has 5700 lines to the centimetre, giving $e = \frac{1}{5700}{}^{-1}$ cm. For the D_1 line of wave-length $5{\cdot}896 \times 10^{-5}$ cm., the angle of diffraction for the first order is given by

$$\frac{1}{5700} \sin \theta = 5{\cdot}896 \times 10^{-5}$$

whence $\theta = 19° \ 52'$. Similarly the first order angle of diffraction for the D_2 line of wave-length $5{\cdot}890 \times 10^{-5}$ cm. is $19° \ 36'$. The corresponding angles for the second order are $42° \ 49'$ and $42° \ 10'$. The third orders have values of $\sin \theta$ just greater than 1, and therefore are not produced with a grating whose element is as small as this.

This is quite a common value for the grating element and the reader may feel tempted to ask why it is chosen to give so few orders of spectra. One reason is that the greater the angle of diffraction the more accurately it and the wave-length of light may be measured. Another reason may be this. The second order violet and the first order red are produced in directions given by the equations $e \sin \theta_2 = 2\lambda_v$ and $e \sin \theta_1 = \lambda_r$, in which the notation explains itself. Since $\lambda_r = 2\lambda_v$, $\theta_1 = \theta_2$, so that the second order violet begins just about where the first order red stops. Again $e \sin \theta_3 = 3\lambda_v$ and $e \sin \theta_2 = 2\lambda$. If $\theta_3 = \theta_2$, $\lambda = \frac{3\lambda_v}{2}$ which is a wave-length in the yellow. Therefore the third order violet begins at the yellow of the second order spectrum. In fact, the second and third order spectra overlap a little and the reader can easily show for himself that this overlapping gets worse as we go to higher orders. This overlapping might

lead to confusion in the case of line spectra with many lines and so the ordinary grating is arranged to give only two orders.

Since most of the light goes straight on through the grating, the diffracted images will not be very bright and, in the case of faint sources of light, it may be difficult to see them and to measure the wave-length of the lines of the source accurately. And the lines get fainter as we go to higher orders. The reader may be inclined to reply: Why not use brighter sources of light then? The answer is that the brightness of the source is often not under our control, as, for example, when we are trying to measure the wave-length of the lines of the spectrum of the aurora borealis in order to find out more about its nature. So it may be desirable to make a grating in which all the diffracted light is concentrated into one order, say the first order. It turns out that this can be done, in theory at any rate. A grating of such dimensions is constructed that it produces only two orders and the second order is suppressed in the following way The grating consists of a set of individual slits each of which has its diffraction pattern consisting of a central maximum and a set of subsidiary maxima separated by minima. It is arranged that the direction of one of these minima coincides with that of the second order maximum for the grating as a whole; therefore; although the beams from the slits reinforce one another, the intensity of each beam is zero and the second order spectrum has zero intensity. It has been suppressed! The above coincidence is achieved in the following way. The direction of the second order maximum for the grating as a whole is given by

$$(a+b) \sin \theta = 2\lambda$$

where a is the width of a slit and b is the width of the opaque line separating one slit from the next. If the first minimum for the diffraction pattern of each slit considered as an individual lies in the same direction, we have

$$a \sin \theta = \lambda$$

Dividing the first of these two equations by the second, we have

$$\frac{a+b}{a} = 2$$

or

$$a = b$$

So it is only necessary to arrange for the slits and the opaque lines to be of the same width to concentrate all the light into the first order spectrum and to make it as bright as possible.

We have already seen that the spectrum produced by a diffraction grating differs from that of a prism in that the violet is the least deviated and we shall conclude this account of the grating by drawing attention to another difference. The dispersive power of a grating is defined as the rate at which the deviation varies with the wave-length, or as $\frac{d\theta}{d\lambda}$.

Now we have

$$e \sin \theta = n\lambda$$

Differentiating this equation, we have

$$\frac{d\theta}{d\lambda} = \frac{n}{e \cos \theta}$$

So the dispersive power increases as the order of the spectrum increases and as the number of lines per centimetre of the grating increases. The dispersive power tells us the angle between the parallel beams of two neighbouring wave-lengths, or the distance apart of their images in the focal plane of the objective of the telescope of the spectrometer. If a spectrum of white light is produced, the value of $\cos \theta$ alters so little between the two ends of the spectrum that it can be treated as constant; so the distance between two lines of a given difference in wave-length is the same all the way along the spectrum. This is called a normal spectrum and, if it is compared with that produced by a prism, it will be seen that the red and yellow occupy a greater length and the blue and violet a shorter length than in the spectrum of a prism. The reason for this is that the dispersive power of a prism is greater at the violet than at the red end of the spectrum, thus causing the blue and violet to be unduly spread out or dispersed. We shall see that it is important to have a normal spectrum when quantitative measurements on the energy in spectra are being made.

143. RESOLVING POWER

What is it that we really expect of an optical instrument such as a telescope or a spectrometer? We expect a telescope to give us a bright magnified image of the moon, for example, although we have seen that the image can never be brighter than the object. We expect a spectrometer to spread out the light from the source to be analysed into a band, in which each individual colour is separated from its neighbours. But are we satisfied by mere magnification in the case of the telescope or microscope, for example? Would the telescope be any use if it produced an image ten times as big as that which we see with the naked eye but showing no more *detail*, no more *grain*? If we pointed it at the moon, all we would see would be an enlarged view of the man in the moon! It would be very much like what we should see if an artist painted a picture of the moon on an elastic canvas and we pulled it out tenfold in each direction. This is no use at all. We do not want just an enlarged view of the moon; we want to see as separate objects the many craters which merge into one when seen by the naked eye. Fortunately the telescope *does* reveal more grain when it magnifies, but it seems to have happened largely by accident! **The amount of detail or grain revealed in an object by an optical instrument is called its resolving**

power. Increase in resolving power seems to accompany an increase in magnifying power, but we shall be wise to investigate the factors which control the resolving power of optical instruments to see if their performance can be improved.

The quantitative definition of resolving power naturally depends on the purpose of the instrument and so varies from instrument to instrument. To begin with the telescope, its purpose is to view distant objects and the amount of grain or detail which it reveals depends on how small an angle two objects may subtend at the objective and yet be distinguished as separate. So **the resolving power of a telescope is defined as the angle subtended at its objective by two point objects which can just be distinguished as separate.**

The reader may now be tempted to ask, what is the difficulty? A telescope produces a point image of a point object; therefore, if two points are separate they must produce separate images! But we must remind him that point images are never produced when paraxial conditions are violated; we have to put up with the circle of least confusion

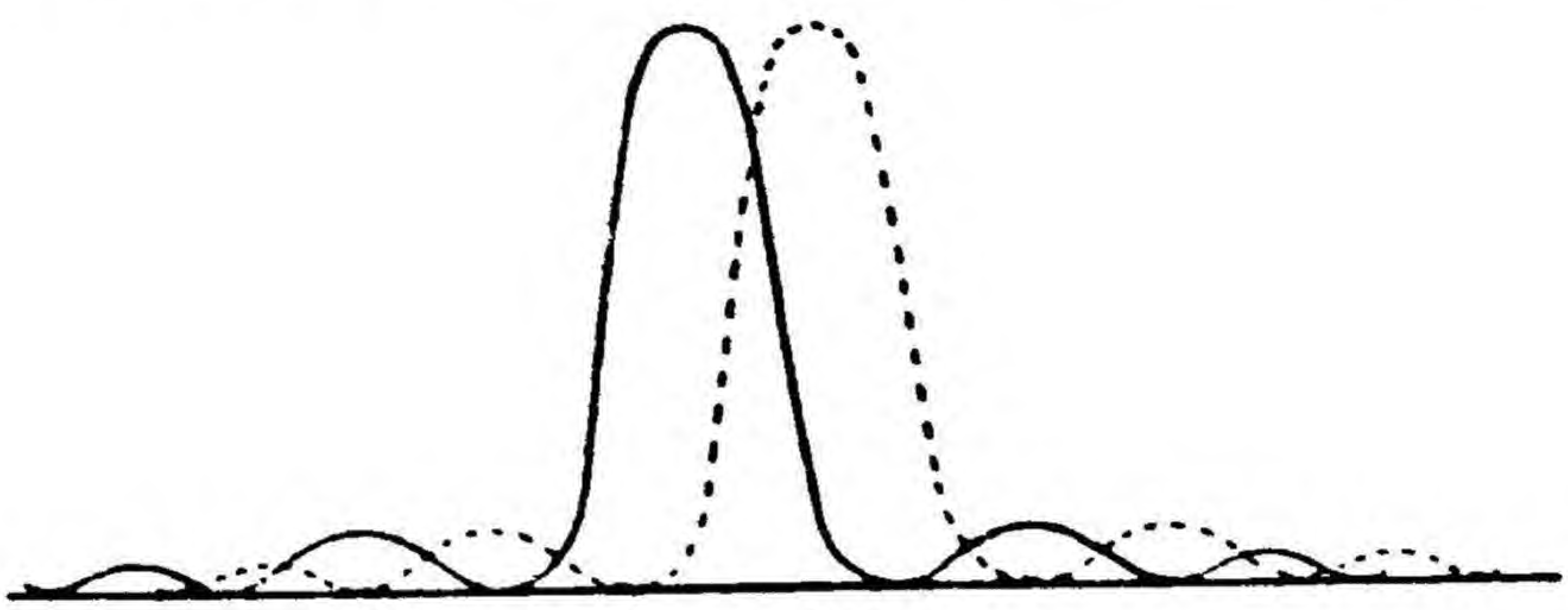


Fig. 228.

due to the various lens aberrations. But this circle of least confusion has been reduced to such a small one that the eye itself cannot distinguish it from a point. Is this not sufficient? But this argument is based on the laws of geometrical optics and on rectilinear propagation, which we have proved to be only approximately true. Light does not travel quite in straight lines and, when a beam of light passes through a lens, the outline of the lens acts as a diffracting system and a **diffraction pattern** of this outline is produced instead of the point image or circle of least confusion. If the outline of the lens is circular, the diffraction pattern is a bright disc, surrounded by alternate dark and bright rings (Art. 136), while, if the lens is cylindrical, its outline is the same as a slit, whose diffraction pattern is a bright central maximum with subsidiary maxima separated by minima on either side. Each distant point object will produce its own diffraction pattern with the centre at the place where the point image would occur. The question we have to decide is *how far apart must these diffraction patterns be for the two patterns to be distinguished as separate?* It is clear that they must be far enough apart to allow the intensity in between the two centres to diminish by a finite amount, but

the actual criterion has to be settled experimentally by placing two objects at such a distance apart that they can only just be distinguished by the eye and then calculating the separation of their diffraction patterns. It turns out that the average observer can separate two images, when the maximum of one pattern falls on the first minimum of the other as shown in Fig. 228.

We can now calculate the resolving power of a telescope with a cylin-

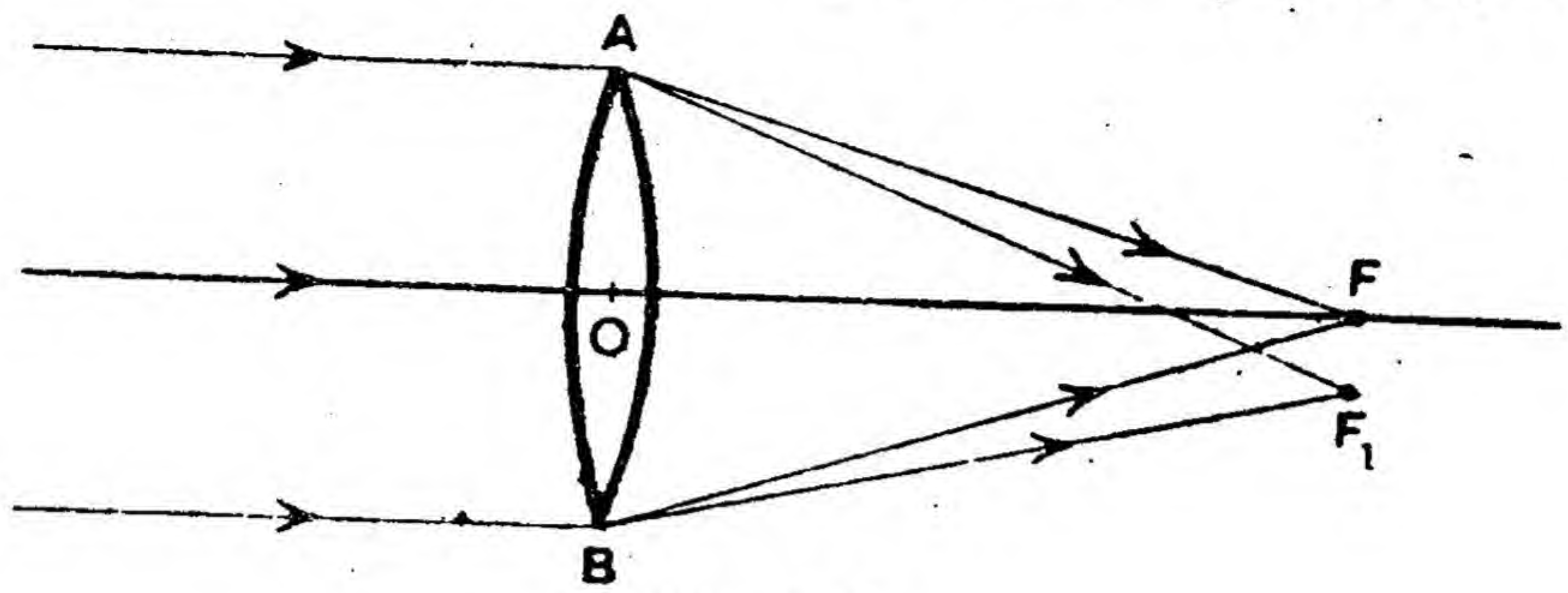


Fig. 229.

drical objective (Fig. 229) bringing a parallel beam of rays from a distant object to a line focus at F. The first minimum of the diffraction pattern produced by the outline of the lens will be at F_1, such that $AF_1 - BF_1 = \lambda$. Then a wavelet from A destroys one from O at F_1 and similarly for any pair of wavelets on the plane wave front at the lens half the width of the lens apart. Hence the wave front can be divided into two halves AO and BO, the one destroying the other at F_1 by interference. By the same reasoning as was used in Art. 121

$$AF_1 - BF_1 = \frac{AB \times FF_1}{OF}$$

$$\therefore \quad FF_1 = \frac{\lambda f}{D}$$

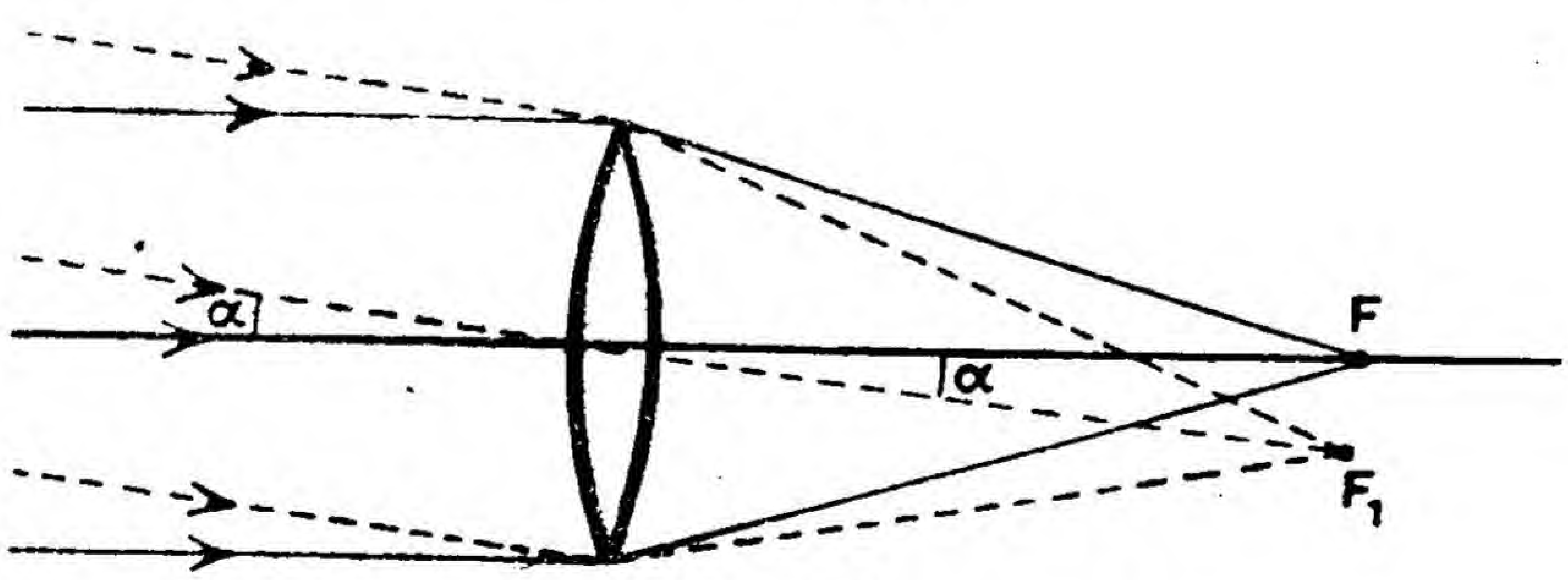


Fig. 230.

where D is the width AB of the lens. This object, at which the axis of the telescope is pointed, will be resolved from a second object if the centre of the diffraction pattern of the second object, or its geometrical image, is at F_1. Its rays therefore make an angle α with those from the first object (Fig. 230), so that α is the resolving power of the telescope. But

$$\alpha = \frac{FF_1}{f} = \frac{\lambda}{D}$$

If the telescopic objective is a spherical lens with a circular outline, the diffraction pattern is a central disc surrounded by a set of alternate dark and bright rings. Strict mathematical treatment shows that the radius of the first dark ring is $\frac{1 \cdot 22\lambda f}{D}$ and, as before, two objects will be resolved when the centre of the diffraction pattern of one falls on the first dark ring of the other. So the angle subtended at the objective by two objects which can just be distinguished as separate is $\frac{1 \cdot 22\lambda}{D}$, which is the resolving power of the telescope. Taking λ as 6×10^{-5} cm., the resolving power of the 40-in. Yerkes refracting telescope is $\frac{1 \cdot 22 \times 6 \times 10^{-5}}{40 \times 2 \cdot 54}$ radians or $0 \cdot 15''$. The resolving power of the eye is about $2'$, and so the telescope must magnify $\frac{2 \times 60}{0 \cdot 15}$ or about 800 times to take full advantage of the large diameter of the objective. But any further magnifying power is useless, since it would not be accompanied by any more resolution. This point must always be taken into account in the design of a telescope.

144. THE DIFFRACTION GRATING

The function of a diffraction grating is to produce a spectrum and grain in a spectrum means the ability to separate two lines whose wave-lengths are very nearly the same, such as the sodium D_1 and D_2 lines. The less the difference in wave-length between two lines which can just be separated, the greater the resolving power of the grating and it is defined quantitatively as **the ratio $\frac{\lambda}{d\lambda}$, where $d\lambda$ is the difference in wave-length of two lines of mean wave-length λ which can just be separated.**

A full mathematical treatment of the grating shows that the diffraction pattern consists of a set of main maxima, whose directions are given by equation (79), separated by alternate minima and subsidiary maxima. As with the telescope, two lines will be distinguished as separate by the average observer when the principal maximum of one falls on the first minimum of the other. It seems, then, as if it is quite easy to increase the resolving power of a grating. It is only necessary to make it produce a wider spectrum by increasing its dispersive power and increased resolving power will follow. But dispersive power in a grating is like magnifying power in a telescope and increase in resolving power does not accompany increase in dispersive power any more than it does increase in magnifying power. For, if a grating produces the diffraction patterns of the two sodium lines with their principal maxima at A and B (Fig. 231), the lines will not be resolved, because the principal maxima are too wide. If

another grating of greater dispersive power is used, it will make the centres of the diffraction patterns at A′ and B′ further apart, but this will be no use if it widens each maximum in the same ratio! Therefore the

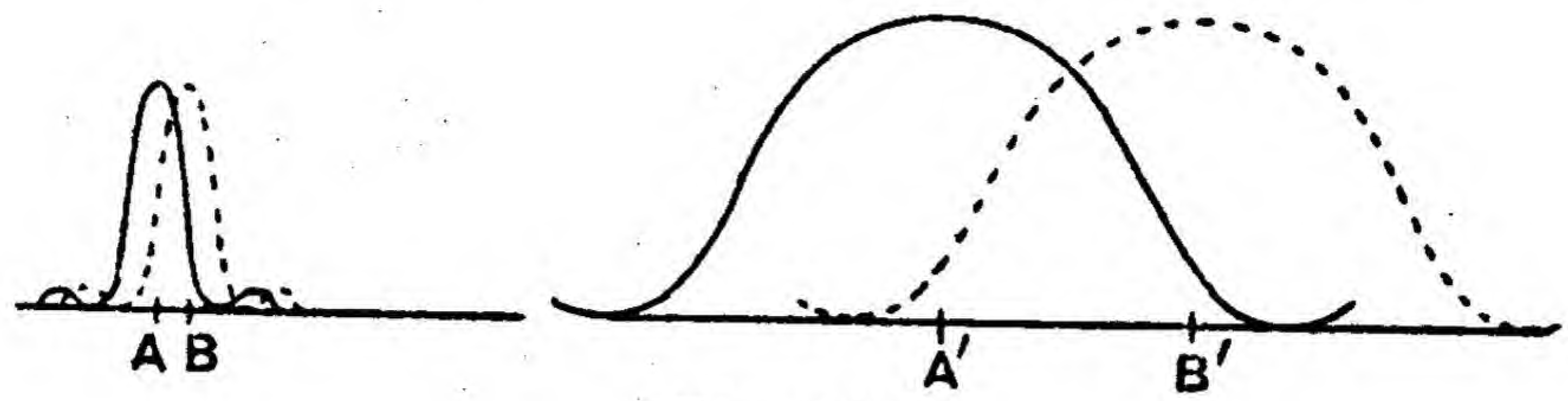


Fig. 231.

resolving power of the grating can only be increased by making the central maxima narrower without altering the dispersion of the grating. We must therefore calculate the angular separation of the centre of the principal maximum and the first minimum. Let AB represent a horizontal cross-section of the complete grating with N lines altogether and let a parallel beam of monochromatic light of wave-length λ fall normally on it (Fig. 232). Let N be an even number. Then, if the path difference between the rays from the extreme slits at A and B diffracted in a certain direction is λ, there will be a minimum in that direction. This is because the path difference between the rays from any two slits half the width of the grating apart is $\frac{\lambda}{2}$ and so the two rays destroy each other. Hence the rays from one half of the grating destroy those from the other half and so a minimum is produced in the given direction. Minima are also obtained in the directions for which the path difference between the rays from A and B is 2λ, 3λ, 4λ, and so on. The minima for path differences of 3λ, 5λ, and so on, can be established by dividing the grating into two halves; that for path differences of 2λ, 4λ, and so on, by dividing it into 4, 8 parts and so on. If N is an odd number, the above argument is applied to N$-$1 lines of the grating, the light from the remaining slit being ignored, as its intensity is too small to matter. But principal maxima are produced in those directions for which the path difference between the extreme rays from A and B is Nλ, 2Nλ, 3Nλ, and so on, since rays from consecutive lines reinforce one another. Let the direction AC be the principal maximum of the n^{th} order spectrum of wave-length λ. Then

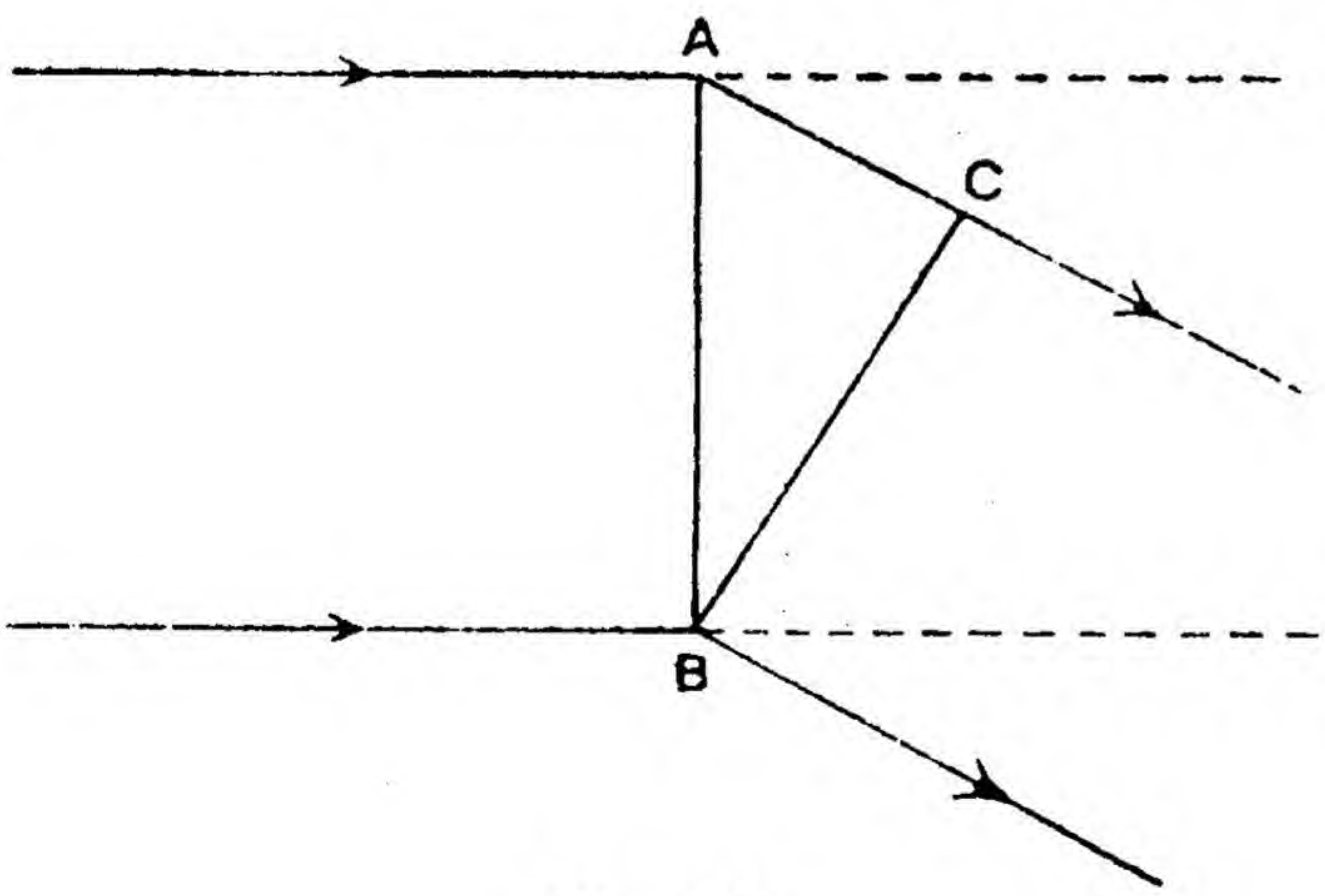


Fig. 232.

$$AC = nN\lambda$$

If a slightly smaller wave-length $\lambda - d\lambda$ is sent normally on to the grating, its principal maximum in the n^{th} order spectrum will lie along a direction slightly less inclined to the direct beam than AC. But, if this line is to be separated from that of wave-length λ, the first minimum *after* the above principal maximum must also lie along AC. Hence

$$AC = (nN+1)(\lambda - d\lambda)$$

$$\therefore nN\lambda = (nN+1)(\lambda - d\lambda)$$

$$\therefore \lambda - nNd\lambda - d\lambda = 0$$

$$\therefore \frac{\lambda}{d\lambda} = nN + 1$$

$$= nN \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (80)$$

Since N may be 10,000 and n may be 2, it is quite legitimate to neglect the 1 in comparison with nN. Therefore the resolving power of the grating is controlled largely by the total number of lines in the grating, while the expression for the dispersive power shows that it is controlled largely by the number of lines per unit length. If this latter number is increased so as to get the centres of two lines further apart without increasing the total number of lines in the grating, the resolving power will not be increased. If the lines were not resolved before, they will not be resolved by the new grating and for precisely the reason illustrated in Fig. 231. It really comes to this: the less the angle between the principal maximum of a line and its first minimum, the greater the resolving power of the grating. In order to go from the maximum to the next minimum, the direction of the rays must be changed so as to increase or decrease the path difference AC by λ. The greater the number of wave-lengths AC contains, the less this change in direction will have to be; that is, the greater the number of lines in the grating, the less the angle between the maximum and minimum and so the greater its resolving power.

145. THE PRISM

Let AE represent a plane wave front of monochromatic light of wave-length λ incident on a prism, which produces a refracted wave front DH. This wave front passes through a slit limiting the amount of the prism effectively employed by the beam of light passing through the slit. If n is the refractive index of the glass for the given wave-length, we have by the principle of equal times

$$AB + n.BC + CD = EF + n.FG + GH$$

If a second plane wave of wave-length $\lambda - d\lambda$ is sent along the same path, the refracted wave front will be DK, the angle between the two wave fronts having been considerably exaggerated. In this case we have

$$AB + (n+dn)BC + CD = EF + (n+dn)FG + GK$$

Subtracting the first equation from the second, we have

$$dn \,.\, BC = dn \,.\, FG - HK$$

$$\therefore\ dn(FG - BC) = HK$$

The principal maximum of the wave-length λ lies along the direction normal to DH, while that of $\lambda - d\lambda$ lies along the normal to DK. If these two lines are to be resolved, the first minimum of λ must lie along the normal to DK. This minimum is produced by diffraction at the slit and the path difference between the extreme rays from D and H in that direction must be λ. This path difference is equal to the length of the normal from H on to DK, which is almost equal to HK.

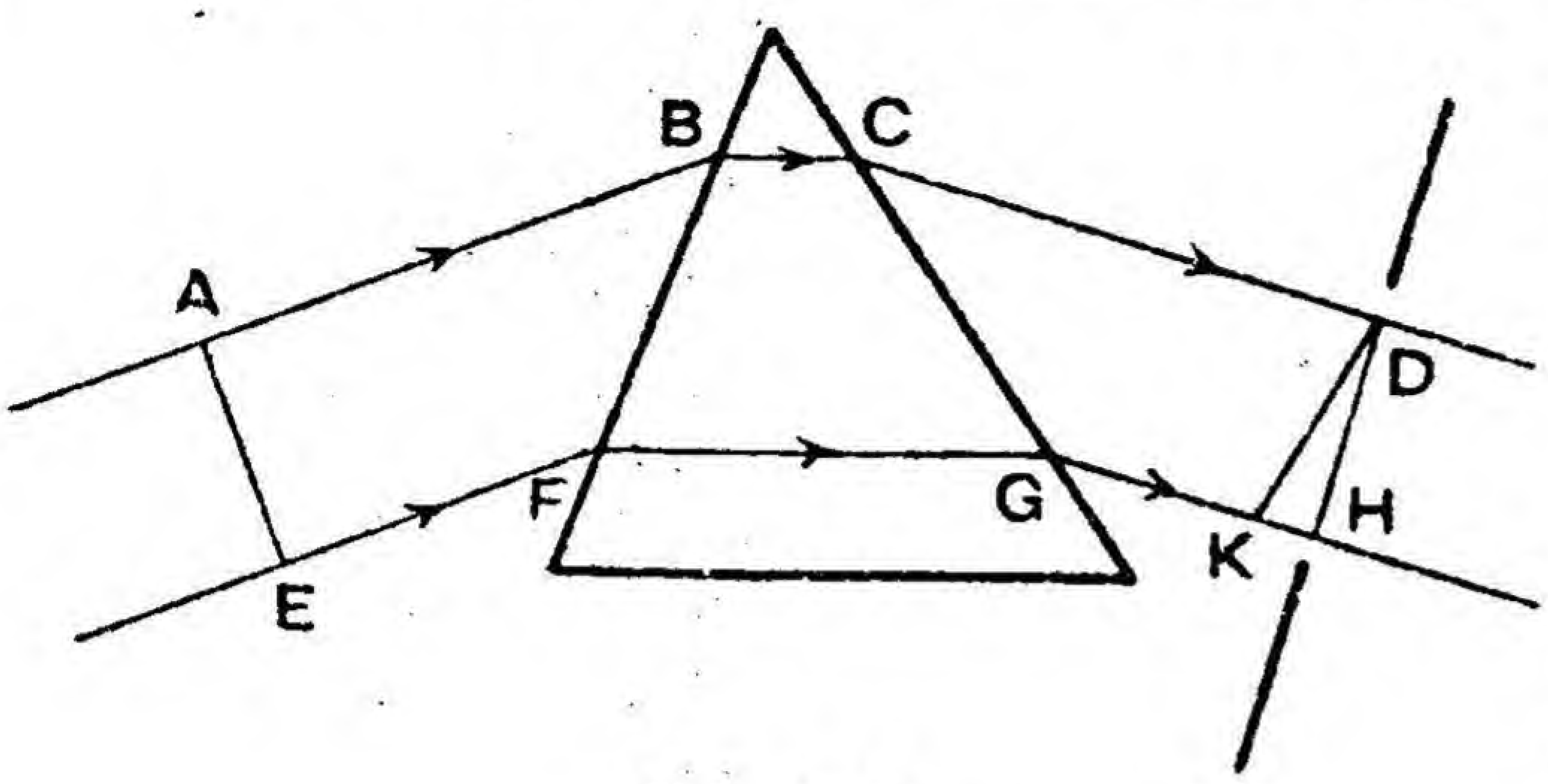


Fig. 233.

$$\therefore\ HK = \lambda$$

$$\therefore\ dn(FG - BC) = \lambda$$

$$\therefore\ \frac{\lambda}{d\lambda} = (FG - BC)\frac{dn}{d\lambda}$$

The greatest value which (FG—BC) can assume is the length, t, of the base of the prism. Therefore the greatest value for the resolving power of the prism is given by

$$\frac{\lambda}{d\lambda} = t\frac{dn}{d\lambda} \quad \ldots \ldots \ldots \quad (81)$$

To resolve the two sodium D lines of wave-length $5{\cdot}896 \times 10^{-5}$ and $5{\cdot}890 \times 10^{-5}$ cm. needs a resolving power of $\dfrac{5{\cdot}896 \times 10^{-5}}{0{\cdot}006 \times 10^{-5}} = 1000$. It is quite easy to get this with an ordinary grating having 5000 lines to the centimetre used with a spectrometer whose telescope will use 1·5 cm. of the grating, giving a resolving power of 15,000 in the second order spectrum. But it is not easy to get the necessary resolving power in the usual flint glass prism.

We started out in this chapter to answer two questions and to follow up a clue. If light is waves, why does it travel in straight lines? And can we demonstrate, in the case of light, the diffraction which other waves show? Are the bands observed by Grimaldi inside and outside the shadow of a wire an example of this diffraction which we are seeking? We have obtained a satisfactory answer to all our questions. Light travels in straight lines only because its wave-length is so small compared to the linear dimensions of the obstacles and apertures used in demonstrating rectilinear propagation; it *can* be diffracted as is shown by the bright spot at the centre of the shadow of a circular obstacle; and

Grimaldi's bands were indeed due to diffraction, as the combination of theoretical prediction and experimental measurement showed. But we have gone a long way beyond our original terms of reference. In studying other cases of diffraction for their own sake, we have been led to the discovery of the diffraction grating, a most powerful weapon for the measurement of wave-lengths. We have hinted at its power to unravel the complexities of the spectral lines and we may expect much from this new field of investigation which opens itself before us. For dare we not hope that the spectral lines hold the key to the nature of the atom, since they are emitted by atoms themselves? We shall make a brief reference to this topic later, but we must close on a more practical note. Our interest in the nature of light has led to a deeper understanding of the purpose of optical instruments and how they achieve greater resolving power; so once again the pursuit of knowledge for its own sake has led to fruitful practical application and the debt which pure knowledge owed to industry, in that the desire to improve telescopes was the starting-point of Newton's work in Optics, may now be said to have been repaid.

EXAMPLES ON CHAPTER XIV

1. Explain how the wave theory of light accounts for reflection, refraction, and diffraction at a slit. Explain how to calculate the wave-length of light by measurement of the diffraction pattern. (*Oxford Schol.*)

2. A parallel beam of white light strikes at right angles a plate in which a very small circular hole is bored and falls on a screen after passing through the hole. Describe the appearance on the screen and how it varies as the distance between the plate and the screen increases and decreases.

If the plate with the hole in it were replaced by a small circular disc, at right angles to the incident light, what would be the appearance on the screen? What part did the latter experiment play in the controversy which arose as to the validity of the theory of interference? (*Camb. Schol.*)

3. Write a short account of the bending of light waves round obstacles, and explain in particular (a) the fringes observed near the boundary of the geometrical shadow of a straight edge, and (b) the bright spot which may be observed at the centre of the shadow of a suitably illuminated disc. (*London B.Sc.*)

4. Discuss the phenomena of diffraction, illustrating your answer by considering in detail one of the following cases:

(a) The distribution of intensity along the axis of a circular aperture illuminated by monochromatic light from a point source on the axis.

(b) The distribution of intensity in the focal plane of a converging lens illuminated by a parallel beam of monochromatic light, the aperture of the lens being limited by a slit. (*Tripos, Part I.*)

5. Discuss the phenomenon of diffraction at a straight edge. Describe fully how you would obtain the wave-length of light in this case, stating precisely what measurements you would make on the apparatus and the fringes and how you would calculate the wave-length of light from them.

6. A vertical slit illuminated with sodium light is placed in front of a vertical screen and in between there is placed (a) a narrow vertical aperture, (b) a narrow

vertical obstacle. Describe fully and explain carefully the effects observed on the screen in each case, making suitable comparisons. What bearing have these effects on the wave theory of light?

7. A very narrow vertical slit illuminated with sodium light of wave-length $5{\cdot}9 \times 10^{-5}$ cm. casts a shadow of a vertical copper wire 1 mm. in diameter and 2 metres away on a screen 3 metres from the wire. Calculate the distance apart of the bands inside the shadow and the total number of bands which can be seen.

8. Describe and explain the diffraction pattern which is seen in the focal plane of a lens of 1 metre focal length placed just behind a slit of width 1 mm. which is illuminated by a parallel beam of monochromatic light.

How would you in practice produce the parallel illuminating beam? Give approximate magnitudes of the important parts of any apparatus which you would use. (*Camb. Schol.*)

9. What is a zone plate? Show clearly how it forms an image of an object, and derive an expression for its focal length in terms of the necessary quantities.

It is required to make a zone plate with a focal length of 50·0 cm. for light of wave-length 6×10^{-5} cm. Concentric circles whose radii are proportional to the square roots of the natural numbers have been drawn on paper, the radius of the smallest circle being 10·0 cm. Calculate the distance at which you would place this paper from a 10·0 cm. focal length converging lens in order that the image formed may be the right size for the above zone plate.

10. Compare the chromatic aberration of a zone plate with that of an ordinary lens. Discuss the possibility of making an achromatic lens of a zone plate and an ordinary crown glass lens.

11. Explain the action of a diffraction grating. Monochromatic light of wave-length $6{\cdot}56 \times 10^{-5}$ cm. falls normally on a grating 2 cm. wide; the first order spectrum is produced at an angle of 18° 14′ from the normal. What is the total number of lines on the grating? (*Oxford Schol.*)

12. A diffraction grating used at normal incidence gives a green line, $\lambda = 5400$ Å.U., in a certain order superimposed on the violet line, $\lambda = 4050$ Å.U., of the next higher order. If the angle of diffraction is 30°, how many lines are there to the centimetre in the grating? (1 Å.U. $= 10^{-8}$ cm.) (*Camb. Schol.*)

13. How would you use a diffraction grating to determine the wave-length of sodium light? A steel scale engraved with lines 1 mm. apart is used as a reflection grating. Calculate the angular separation of the first and second order spectra for sodium light of wave-length 6×10^{-5} cm. if the light is incident at a grazing angle of 1°. (*Camb. Schol.*)

14. Parallel light from a mercury arc is incident on a plane diffraction grating at right angles to the lines of the grating, but making an angle θ with the normal to the grating. The green radiation (5461 Å.U.) is observed in the third order to be transmitted at an angle of 41° 42′ to the normal, while the violet radiation (4358 Å.U.) in the same order suffers 7° 18′ less deviation. How many lines does the grating have to the centimetre and what is the angle of incidence? (1 Å.U. $= 10^{-8}$ cm.) (*Camb. Schol.*)

15. Describe with the necessary theory, how you would use a diffraction grating to measure the wave-length of sodium light.

A grating has 6000 lines to the centimetre. Find the angular separation of the two yellow mercury lines (wave-lengths 5770 and 5791 Angstrom units) in the second order. (*Camb. Schol.*)

16. Explain the action of a diffraction grating, and describe how you would use it to measure wave-lengths.

Why is there a lower limit to the difference of wave-length which can be detected by a grating spectrometer? (*Camb. Schol.*)

17. When a parallel beam of monochromatic light falls normally on a diffraction grating of m lines per cm. several orders of spectra may be observed, provided the wave-length of the light is not too great. What is the limiting value of the wave-length if the nth order spectrum is just observable (deviation $= 90°$)? Show that by a suitable rotation of the grating it is possible, when the latter condition

has been realised, to reduce the deviation of the *n*th order spectrum to 60°. How, and in what circumstances, may this change alter the resolving power of the instrument for light of neighbouring wave-lengths ? (*Camb. Schol.*)

18. How may the wave-length of light be determined ? (*Camb. Schol.*)

19. Explain the formation of spectra by means of a diffraction grating.
What is the highest order spectrum which may be seen with sodium light of wave-length 5×10^{-5} cm., by means of a grating with 3000 lines per cm. ? Calculate the wave-lengths of the colours in the visible spectrum ($3 \cdot 5 \times 10^{-5}$ cm. to 7×10^{-5} cm.) which coincide with the fifth order spectrum of sodium light. (*Camb. Schol.*)

20. Describe the spectrometer, explaining clearly the functions of its constituent parts.
A simple grating spectrometer, the grating of which has 6000 lines to the centimetre, has its telescope replaced by a camera whose lens is of 30 cm. focal length. What will be the linear separation on the plate of the sodium D lines (wave-lengths 5890 Å.U. and 5896 Å.U. respectively) in the second order spectrum ? (*O. and C.*)

21. Describe a method of measuring the wave-length of light. State the observations you would make, and prove any formula you would use. (*O. and C.*)

22. Write as complete an account as you can of the plane transmission grating. Deduce an expression for the wave-length of a spectral line seen in the *n*th order spectrum, assuming the grating to be adjusted for minimum deviation, and calculate the wave-length corresponding to a minimum deviation of 30° in the second order spectrum, assuming 10,000 lines per cm. (*London B.Sc.*)

23. Describe and give the theory of a method of producing a spectrum by means of a diffraction grating. Explain what is meant by diffraction spectra of different orders, and discuss the conditions which would result in the absence of the spectra of even order. (*London B.Sc.*)

24. Describe and explain the difference between the spectra produced by a prism and a diffraction grating.
Show by calculation that the second and third order continuous grating spectra overlap. (The limits of the continuous spectrum may be taken as 4000 Å.U. and 8000 Å.U.) (*N.U.J.B.*)

25. If a vertical sodium flame is examined through a diffraction grating with its lines vertical held close to the eye, the flame itself is seen together with a number of images on either side of it, the images getting fainter the further they are from the bright central image. Explain how these images are produced and calculate the angle between the central image and the one next to it, if the grating has 5000 lines per cm.

26. Describe what you would expect to see if you were to look at a point source of white light through a handkerchief with its threads horizontal and vertical. Make numerical calculations to give some idea of the magnitude of the effects expected. What will happen if the handkerchief is stretched in a horizontal direction ?

27. Explain the action of the diffraction grating, distinguishing between its dispersive and resolving powers. Light is incident normally on a grating $\frac{1}{2}$ cm. wide with 2,500 lines ; find the angular deviations of the two sodium lines in the first order spectrum. Would you expect the two sodium lines to be distinguishable ? The wave-lengths are $5890 \cdot 2 \times 10^{-8}$ cm. and $5896 \cdot 4 \times 10^{-8}$ cm. respectively. (*London B.Sc.*)

28. Explain the action of a grating spectrometer.
What is meant by the resolving power of a spectrometer ? Deduce an expression for the resolving power of a grating spectrometer. (*Tripos, Part I.*)

29. What evidence is there in support of the wave theory of light ? (*Oxford Schol.*)

30. What evidence have we for considering light as a transverse wave motion? Explain in detail the rectilinear propagation of light on the wave theory. (*Camb. Schol.*)

31. Give an account of the various experiments which have been carried out to discriminate between the corpuscular and undulatory theories of light. (*Camb. Schol.*)

32. How is the rectilinear propagation of light explained on the wave theory? (*Tripos, Part I.*)

33. Why is it possible to do a considerable part of the theory of lenses and of optical instruments generally, using a ray theory rather than a wave theory of light propagation? Under what circumstances would the neglect of the wave theory introduce serious errors? Explain why the image of a star appears to be smaller the larger the objective of the telescope through which it is viewed. (*Camb. Schol.*)

Chapter XV

POLARISATION AND DOUBLE REFRACTION

146. INTRODUCTORY

We have seen that the evidence in favour of the wave theory of light is overwhelming and the next question which arises is what sort of waves? And are they transverse or longitudinal? That is, are the vibrations of any particle in the medium through which the waves are passing along a line perpendicular to the direction of propagation of the waves or paralled to that direction? We can hardly attempt any serious answer to the first of these two questions in this book beyond suggesting that the waves are the propagation of some condition in the ether which pervades the whole of the universe and whose peculiar properties were mentioned when we were discussing theories of light (Ch. 10). It was their conception of the ether which led Huygens in 1690 and Young and Fresnel some hundred years later to reject the idea of transverse waves. They argued in this way: we know that there cannot be transverse waves in the air or any fluid, for that fact, since it has no shape; such a medium can only transmit compressional waves, which are longitudinal. Then how can the ether transmit transverse waves, when it is much less dense than air and much more mobile, since it offers no resistance to the motion of the planets through it? But is this a sound attitude of mind towards this problem? Is it not more scientific to see what facts there may be, which bear on the problem, and let *them* decide the theory? The impossibility of having transverse waves in this ether is not a conclusive argument against transverse waves; it may be an argument against regarding the ether as a very rarefied fluid.

It is interesting that the facts were already known to Huygens and still more familiar to Young and Fresnel, but it was a long time before they would face their natural interpretation. Let us consider them. Calcite or Iceland Spar is a mineral form of calcium carbonate and it crystallises in the form of a rhombohedron, ABCDEFGH (Fig. 234), in which the corners at A and G contain three obtuse angles, while the other corners

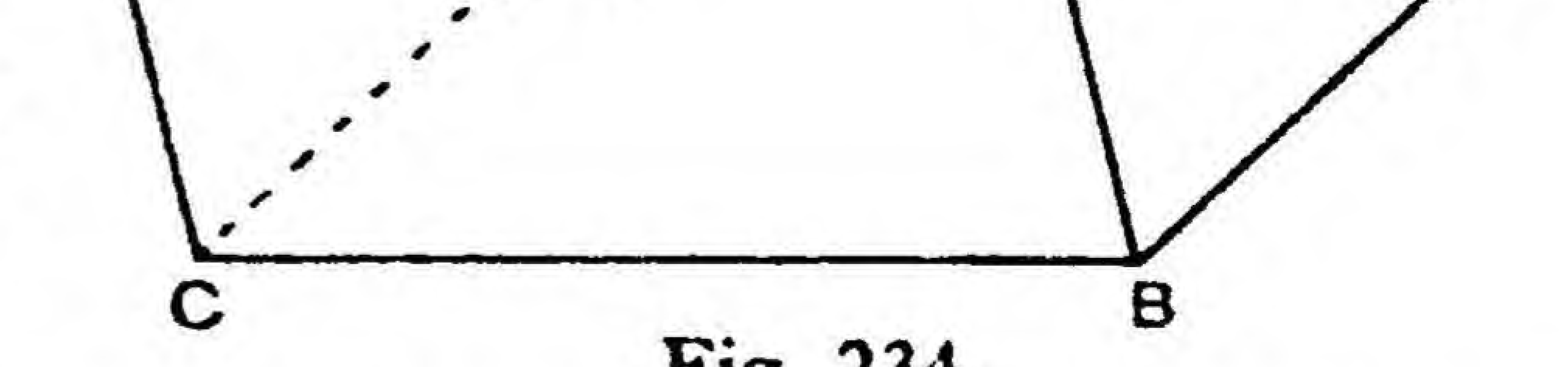


Fig. 234.

contain both obtuse and acute angles. A line making equal angles with the three edges meeting at A, or any line parallel to it, is called an **optic axis** of the crystal; the plane containing an optic axis and the normal to a surface on which light is incident is called a **principal plane.** It should be emphasised that the optic axis is only a *direction*, not a definite line, and the principal plane is not a definite plane; there may be a number of principal planes parallel to a given particular one. If a ray of light falls normally on the surface of a calcite crystal, it is split up into two rays (Fig. 235), one, called the **ordinary** ray, since it obeys the usual laws of refraction and goes straight through the crystal, the other being called the **extraordinary** ray, since it does not obey Snell's law. This is called double refraction and it follows that, if a dot be made on a sheet of paper and a crystal of calcite be laid on it, two images of the dot will be seen (Fig. 236), the one O_1 being due to the ordinary rays, the other E_1 being produced by the extraordinary rays. The extraordinary image

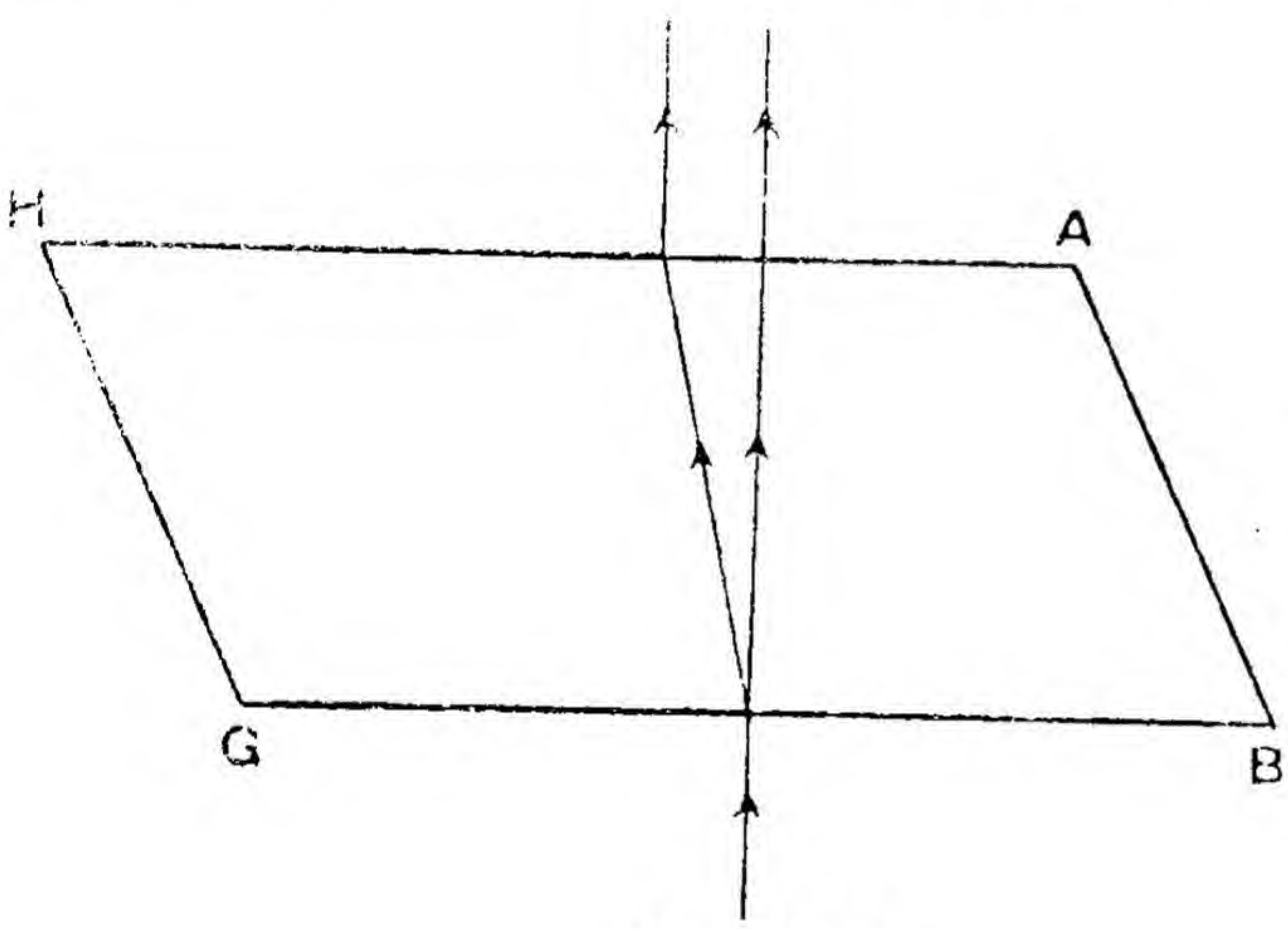


Fig. 235.

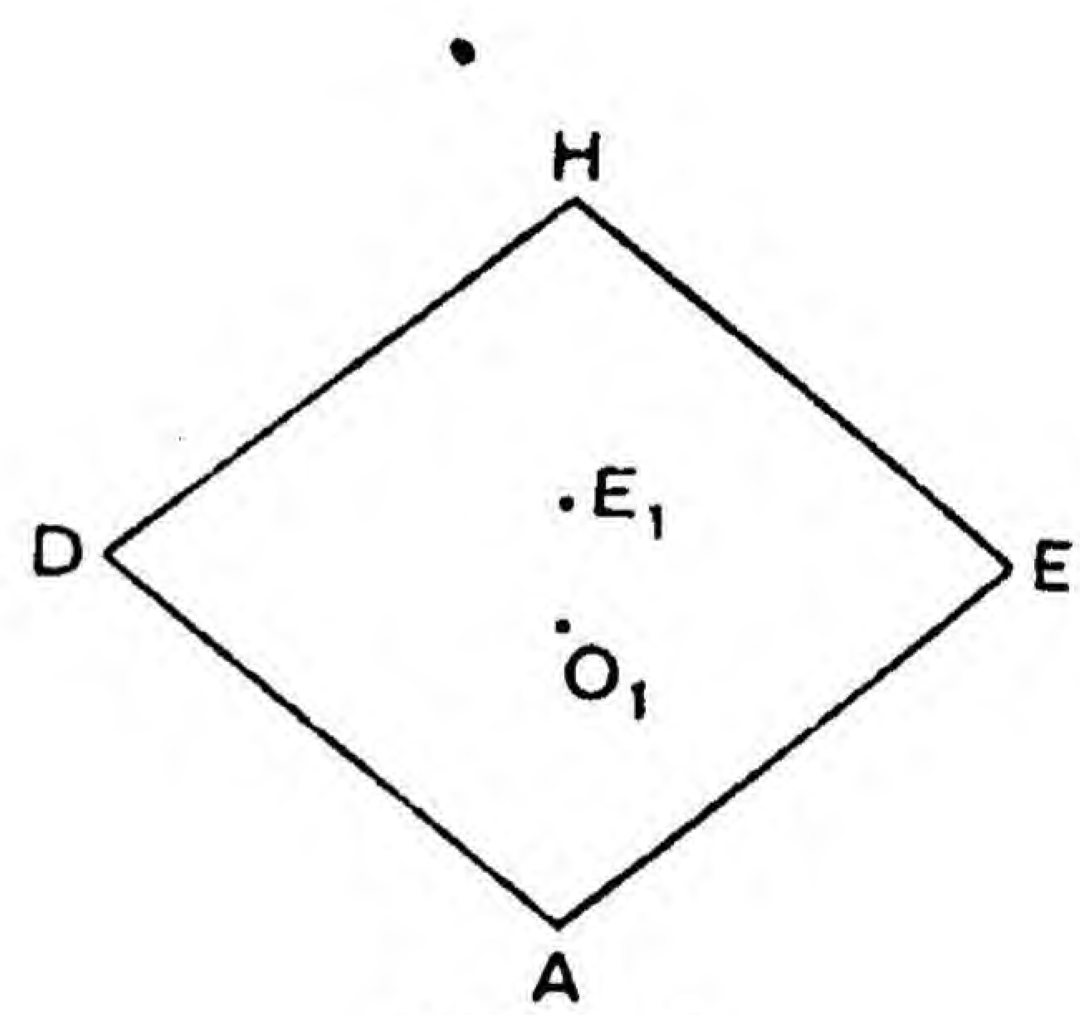


Fig. 236.

is displaced along the diagonal AH starting from one of the two corners containing three obtuse angles and in the sense away from that corner. The amount of the displacement is proportional to the thickness of the crystal. The facts of real interest for the nature of light waves emerge when a second crystal, represented in dotted lines, is placed on top of the first one and rotated (Fig. 237). When the two crystals are in identical positions, two images are still produced, the distance between them being twice as great as before; in other words, the two crystals behave like a single crystal twice as thick as the first one. Another interpretation of this result is that the ordinary ray produced in the first crystal passes through the second one as an ordinary ray without any further splitting up, the same thing being true of the extraordinary ray. If the second crystal is rotated through a small angle about the normal to either refracting surface as axis, two further images are produced in the positions shown; it is clear how they are formed. O_{12} is that due to the ray which goes through both crystals as an ordinary ray, E_{12} that due to the ray going through both crystals as an extraordinary ray, while E_1O_2 is due to the

ray going through the first crystal as an extraordinary ray and the second one as an ordinary ray, and O_1E_2 to that going through the first crystal as an ordinary ray and the second one as an extraordinary ray. This deduction follows from the fact observed with one crystal, that the extraordinary image is displaced along the diagonal starting from one of the corners containing three obtuse angles. The separation between these

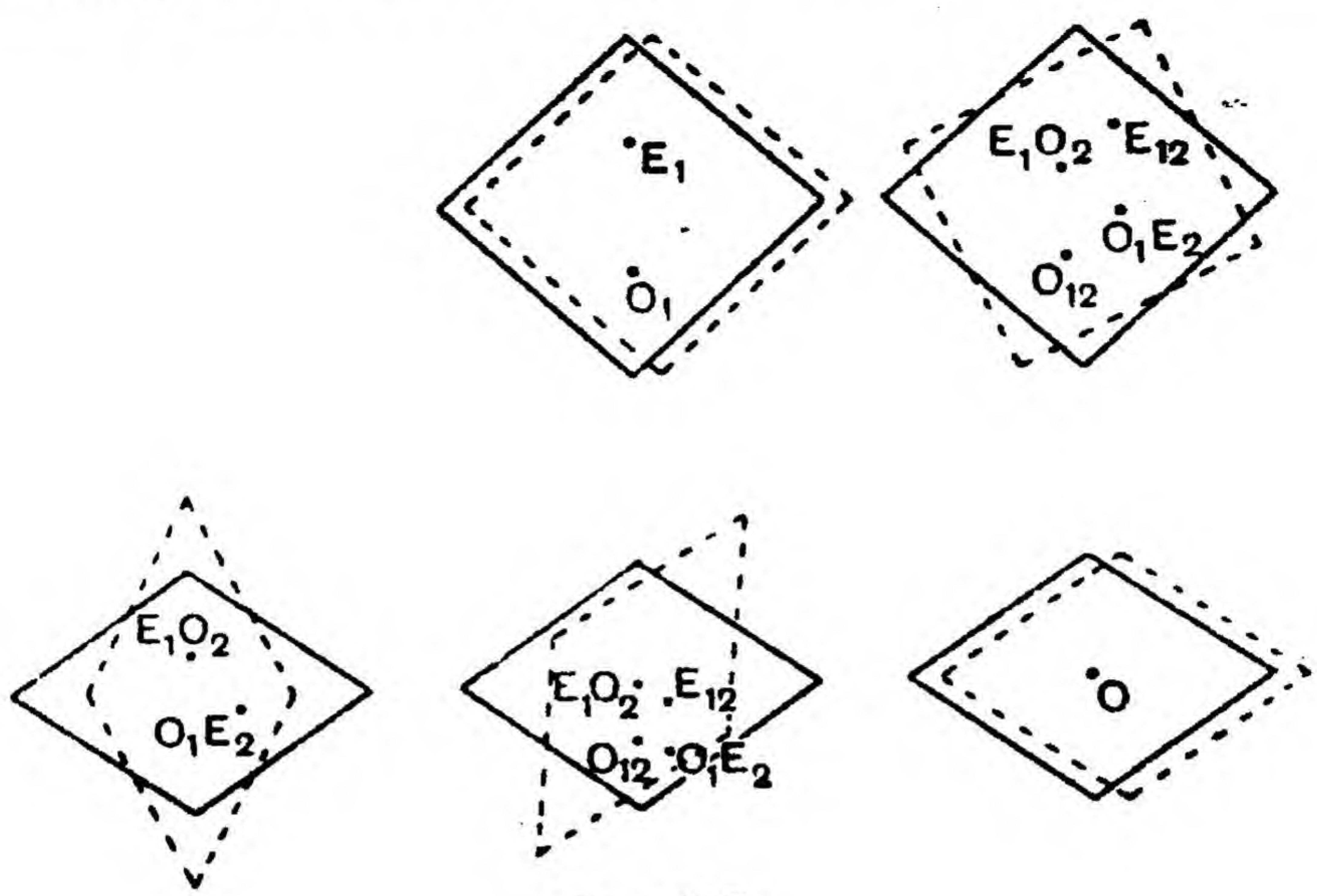


Fig. 237.

pairs of images increases as the second crystal is rotated, but, when it has been turned through a right angle, only two images are formed again, O_1E_2 and E_1O_2. It is clear that the ordinary ray in the first crystal becomes the extraordinary ray in the second and vice versa. A further rotation of the second crystal causes two other images to appear once more, but, when it has been rotated through two right angles, only one image, O, in the position of the ordinary image formed in one crystal is seen. This is due to the fact that the ordinary ray in the first crystal goes through the second as an ordinary ray, while the extraordinary ray in the first goes through the second as an extraordinary ray. As this crystal has been turned through two right angles, its displacement of the extraordinary image is in the opposite sense to that in the first crystal and the image formed by the extraordinary image coincides with the ordinary image.

What are we to make of these facts ? The first crystal splits the light into an ordinary and extraordinary ray and, in general, the second crystal takes each ray and splits it up again into one which behaves as an ordinary ray and the other as an extraordinary ray. The two exceptions to this generalisation are when the crystals are either parallel to or at right angles to each other. What explanation can we offer of this double refraction if light waves are longitudinal ? What is the difference between the ordinary and the extraordinary rays ? Huygens assumed that the ordinary ray travelled in the ether alone, while the extraordinary ray travelled in a medium which was a mixture of the ether and the crystal, the two different media accounting for the difference in behaviour of the two rays. But he

could not give a rational explanation of the above facts with two crystals along these lines. Newton, however, did not allow himself to be obsessed by the impossibility of transverse waves ; with that genius which was his outstanding quality, he allowed the facts, and the facts alone, to dictate his views and he saw clearly from these facts that the only difference in character between the ordinary and extraordinary ray was due to some property which could be expressed as a right angle. This follows from the fact that the ordinary ray in one crystal becomes the extraordinary ray in the other, when the two crystals are at right angles. He also saw that the doubling of the ordinary image by the addition of a second crystal and its becoming single again twice a revolution suggested that the light emerging from the crystal was " one-sided." Expressing his thoughts in terms of the corpuscular theory, he postulated that the corpuscles were to be regarded as having four faces parallel to the direction of the ray of light, one pair having one kind of polarity, something like the N pole of a magnet, and the other pair, at right angles, having the opposite kind of polarity. The faces with one polarity in the ordinary ray were at right angles to those having the same polarity in the extraordinary ray. This would explain, at any rate, the facts observed when the two crystals were either parallel or at right angles to each other. Newton's solution of this problem is yet another example of his uncanny intuition for being guided in his ideas solely by the facts and not allowing himself to be distracted by the apparent impossibility of finding a mechanical model or working thought model to explain his classification of them. His idea of one-sided corpuscles may have been wrong in the light of later evidence, but his conception of the light emerging from the crystal being one-sided was essentially right and the name **polarised light** was derived from his conception of the two sets of poles of the one-sided corpuscles. But it was over a hundred and fifty years before his ideas were incorporated into the wave theory of light ; let us look at the further evidence which was needed to clear up these puzzling facts and to fit them into a consistent scheme.

147. POLARISED LIGHT

We shall start with a simple experiment which can be done with a thin plate cut from a crystal of tourmaline, which is a doubly refracting

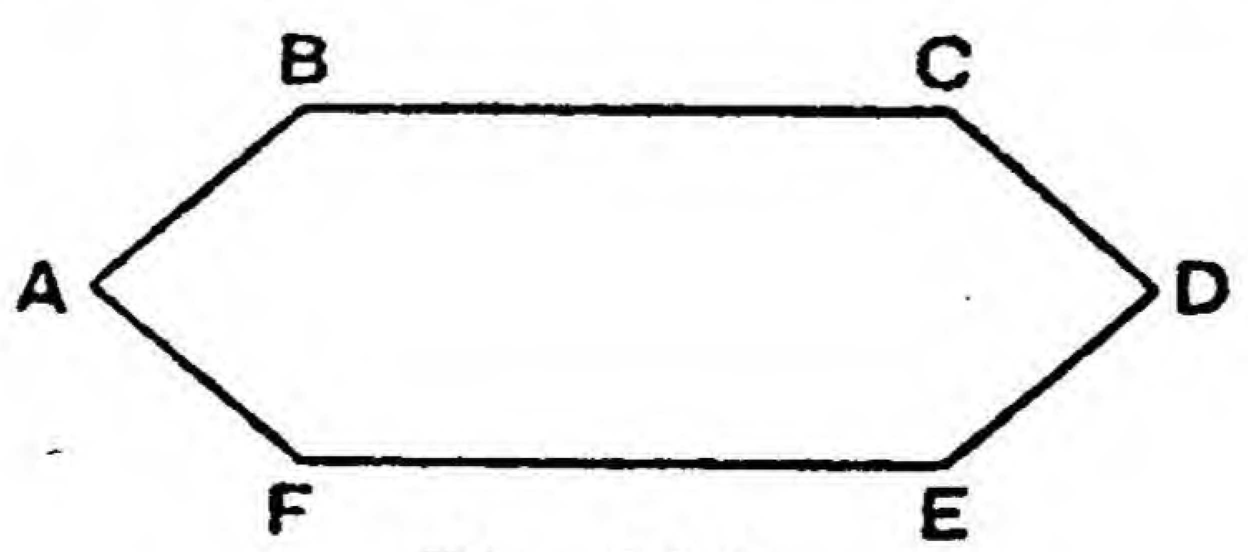


Fig. 238.

mineral crystallising in the hexagonal system. Let ABCDEF (Fig. 238) represent a thin plate of tourmaline cut parallel to a hexagonal base of

the crystal, AD being the principal crystallographic axis of the crystal. Let a ray of light be incident normally on the plate and let it then fall normally on another plate of tourmaline placed parallel to the first one and with their axes parallel. Light will emerge from the second plate. If this plate is now rotated about the ray of light as axis, the intensity of the emergent light will diminish until it is zero when the principal crystallographic axes of the two crystals are at right angles to each other. When the second crystal is rotated further, the intensity of the emergent light increases again until it is a maximum after rotation through a further right angle, when the axes of the two crystals are once more parallel. If the rotation of the second crystal is continued, the emergent light goes through one more minimum and becomes a maximum again when the second crystal has been rotated through four right angles. This is a clear indication of the " one-sidedness " of light and it strongly suggests that the light waves must be transverse, as the following similar experiment with material waves will show. Let a spiral spring and a rope hang vertically from the ceiling side by side and let there be a slot whose width is equal to the thickness of the rope or spring. If compressions and rarefactions are sent along the spring by moving the bottom of it up and down, they will travel through the slot unhindered, if the slot is placed half-way up the spring with its length at right angles to the length of the spring. This is true however much the slot be rotated in its own plane. Now let the slot be placed half-way up the rope in a horizontal position with its length north and south. If the end of the rope be moved to and fro in a north and south direction so as to produce transverse waves, they will pass freely through the slot. Let the slot be now rotated in its own plane, that is, keeping it horizontal. The amplitude of the waves transmitted by the slot will decrease until it is zero when the length of the slot is east and west. It will increase to a maximum again when the slot has been turned through a further right angle, and will go through another minimum and maximum in a further two right angles of rotation. The result of this experiment is precisely similar to that with light and the tourmaline crystals and shows that light waves must be wholly transverse, since the intensity of light passing through crossed tourmalines is zero. If there were any longitudinal component in light, it would get through the tourmaline, which acts as a slot for light waves, and none gets through. This simple experiment, admitting of only one interpretation, forces us to the conclusion that **light waves are transverse**. It is unfortunate that the ether seems to be incapable of transmitting transverse waves, but we shall be wise to postpone this difficulty until further evidence about the nature of light resolves it. There is no question that the above experiment admits of no other interpretation. The history of scientific investigation is full of periods of uncertainty of this kind, in which the facts seem to demand an interpretation which is rationally impossible. Such contradictions are not a sign of failure ; they

are an indication of the existence of a wider truth, which will embrace both sets of ideas and resolve the contradiction.

To return to the above experiment, the first tourmaline plate is called a **polariser**, since it produces one-sided light; it is clear that ordinary light is not one-sided and we shall have to discuss the precise character of the transverse vibrations in ordinary light. The second tourmaline is called an **analyser**, since it is used to detect the presence of the one-sidedness. Both plates are essentially the same; they act as slots for light waves and one is used for producing the one-sidedness and the other for detecting it. Since the vibrations of the particles of the medium produced by the light emerging from the first tourmaline are confined to the same plane, such light is called **plane polarised light.** A mixture of ordinary light and plane polarised light is called **partially plane polarised light**; if such light is examined in the usual way by an analyser, two maxima and two minima are obtained in each revolution of the analyser, but the minima are not complete.

148. POLARISATION BY REFLECTION

In 1808 Malus was doing some work for a **Prize Essay on Double Refraction** and examined the reflection of the setting sun in a window through a calcite crystal. He was surprised to notice that, in certain positions of the calcite, he could see only the ordinary image and for others only the extraordinary image. To see if the effect was due to the reflection at the window or to some effect of the atmosphere on the sun's light, he repeated the experiment with a candle reflected both from a water surface and from glass and found the same results. They suggest that the ordinary and extraordinary rays from a calcite crystal are polarised or one-sided, as Newton had thought over a hundred years earlier, and that the mirror can act as an analyser. If it can act as an analyser, it can serve as a polariser too, and the following experiment can be tried. A ray of light is sent on to a glass plate P (Fig. 239) at an angle i and the reflected ray falls at the same angle of incidence on another plate of glass A, giving rise to a reflected ray parallel to the original incident ray. The plates P and A are set with their planes of incidence parallel and the intensity of the final reflected ray is a maximum. If the plate A is then rotated about the ray incident on it as axis, keeping the angle of incidence the same, the intensity of the final ray goes through two minima and maxima in each revolution, the minima occurring when the planes of incidence of A and P are right angles to each

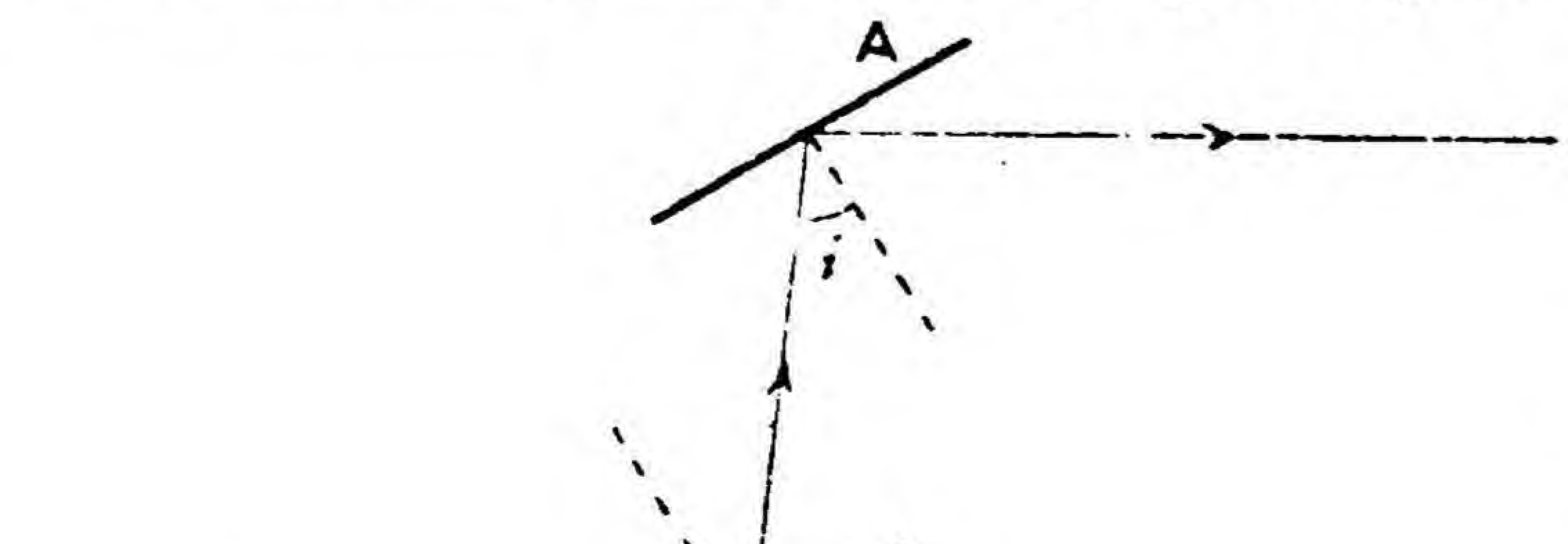


Fig. 239.

other. The minimum intensity of the light is not zero, unless the angle of incidence, i_p, has a special value called the **polarising angle**, which is about 57° for glass. In this case, the first glass plate acts as a polariser and produces plane polarised light and the second plate acts as an analyser. Thus polarised light can be produced by reflection and Brewster found the following relation between the polarising angle and the refractive index n of the reflecting medium :

$$\tan i_p = n$$

which is known as **Brewster's law**. The reader should prove that, when light falls on a mirror at the polarising angle, the reflected and refracted rays are at right angles to one another. It also follows from Brewster's law that, if white light is incident on a mirror, the reflected light can never be quite plane polarised, since the polarising angle is different for different colours and, if the angle of incidence is the polarising angle for one colour, it is not so for any other colour.

When a ray of light falls on a glass plate at the polarising angle, the reflected ray is plane polarised. In what direction are the vibrations of the medium taking place ? To what plane are they confined ? We do not yet know the answer to this question, but until we do we can specify the plane of polarisation by choosing some arbitrary plane which is decided by the position of the mirror producing the polarised light. We may find that the vibrations do not occur in that plane at all, but, as long as they always make the same angle with it, it will suffice for the present. The plane chosen is the plane of reflection of the mirror and it is called the plane of polarisation of the reflected light. So light polarised by reflection is said to be polarised in the plane of reflection. We shall now describe another way of producing polarised light, which tells us the actual plane to which the vibrations are confined.

149. POLARISATION BY SCATTERING

It is well-known that light is scattered when it encounters particles of about the same size as the wave-length of light. The scattering of light by small particles accounts for the blue of the sky and it can be produced artificially by passing a beam of light through a suspension of small particles in water. This can be made by adding a drop or two of a solution of gum mastic in alcohol to a tank of water T (Fig. 240) and, if a horizontal beam of white light is sent through the tank, the suspension will appear to be blue from the side due to the scattered light. If the light scattered horizontally in a direction at right angles to the incident beam is examined in an analyser A, such as a glass plate, it will be found to be plane polarised. This is just what we should expect, since the vibrations in both the incident and scattered beams must be perpendicular to the direction of propagation of the waves and the only direction at right angles to both beams is the vertical direction. Hence the vibrations in the beam scattered

horizontally at right angles to the incident beam must be confined to a vertical plane; thus this beam is plane polarised, as the analyser shows,

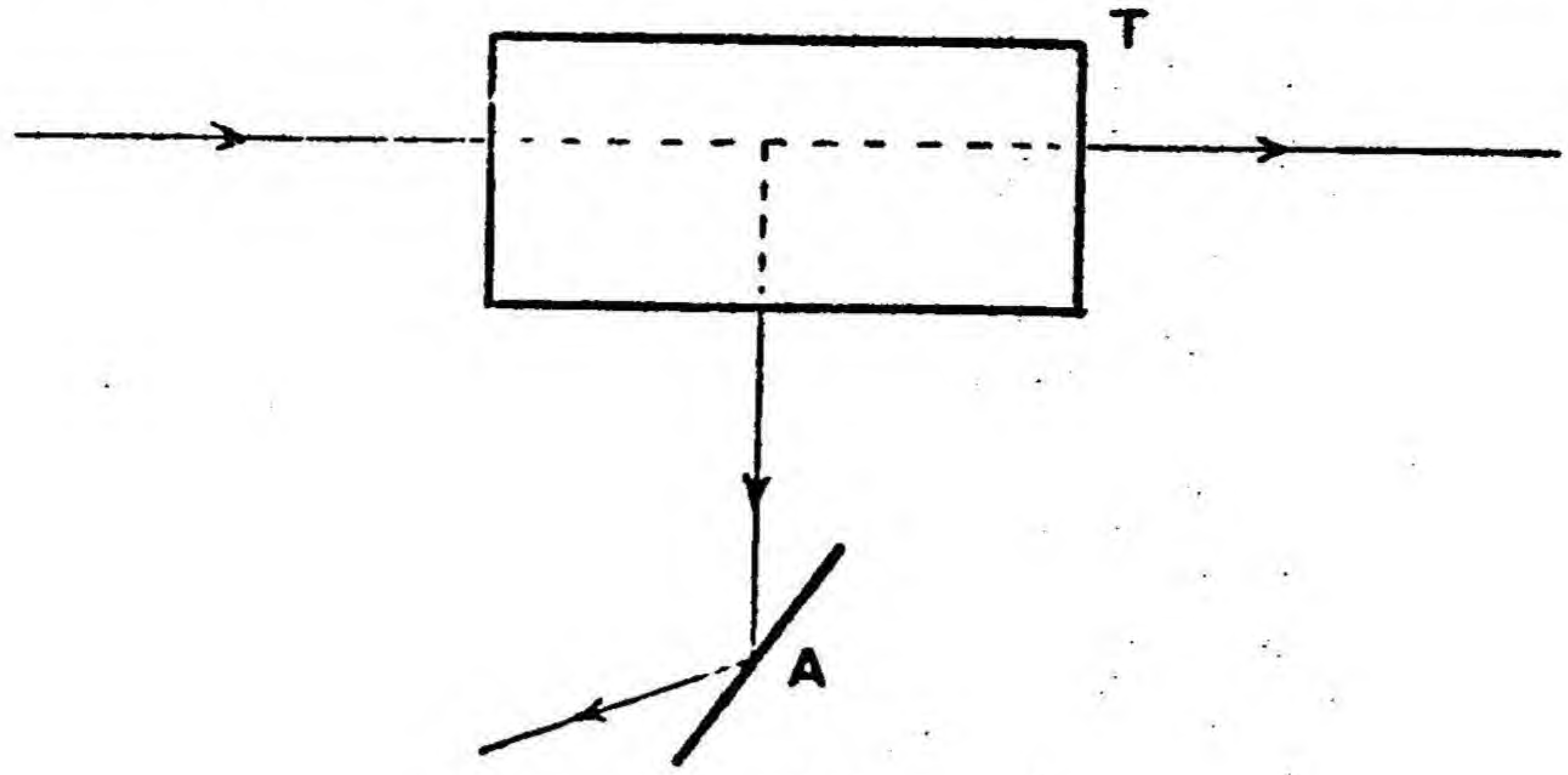


Fig. 240.

and, in addition, *the vibrations are confined to a vertical plane.* But the experiment shows that the light reflected from the analyser is a maximum when the plane of incidence is horizontal, so that the analyser passes vibrations which are normal to its plane of incidence. Therefore it also produces plane polarised light, in which the vibrations are normal to the plane of incidence. The plane to which the vibrations are confined will be called the **vibration plane**, and it is at right angles to the plane of polarisation. In future, we shall specify the direction of vibration in plane polarised light by the vibration plane, since it avoids confusion.

Any analyser acts as a kind of slot for light waves, allowing only those waves to pass whose vibrations are in the direction of the slot. If waves fall on it, whose direction of vibration is not along or normal to the direction of the slot, the analyser allows only that component of the vibration in the direction of the slot to pass. If the vibration plane of the incident light makes an angle θ with the vibration plane passed by the analyser, the amplitude of the vibration passed by the analyser is $a \cos \theta$, where a is the amplitude of the incident wave. Hence the intensity of the transmitted light is $a^2 \cos^2\theta$ or $I \cos^2\theta$, where I is the intensity of the incident light. This is known as Malus's law and has been made the basis of a photometer, since it enables the illumination of a surface to be varied to a known extent.

150. ORDINARY LIGHT

We have now shown that light is a completely transverse wave motion, but we have still to decide the precise nature of ordinary light, which shows no variation in intensity when examined with any known analyser. The analyser will make plane polarised light out of ordinary light; what, then, is ordinary light? For scientific purposes we have to construct a model which will satisfy the above fact about ordinary light, namely, that the intensity passed by an analyser is independent of the position of the analyser about the ray of light as axis. There are two possible

PLATE VI

Fig. 1. Diffraction at a narrow obstacle showing the unequally spaced fringes outside the geometrical shadow and the equally spaced fringes inside it.

(J. W. Cottingham)

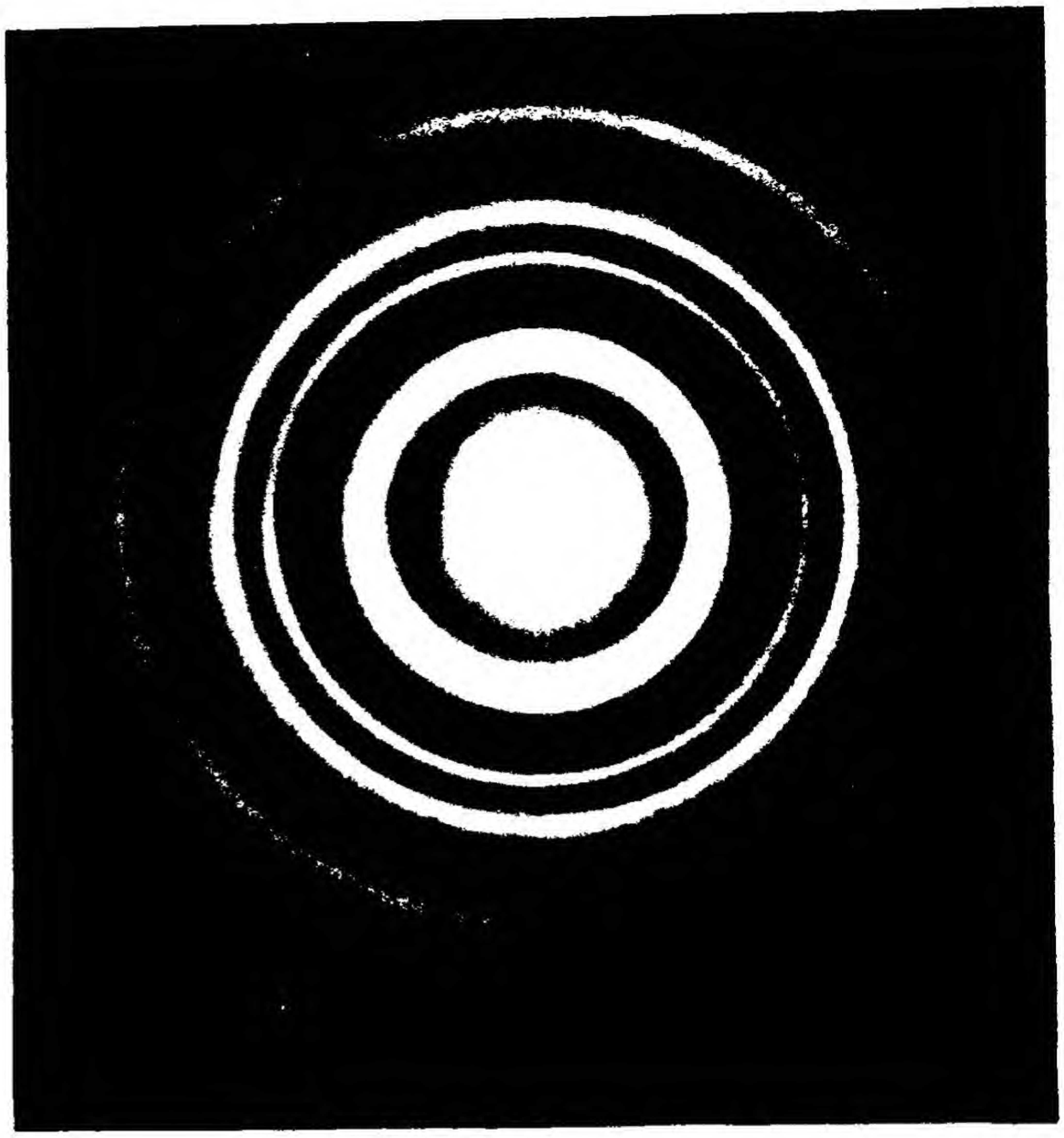


Fig. 2. Diffraction of electrons by a thin film of metal.

(Professor G. P. Thomson)

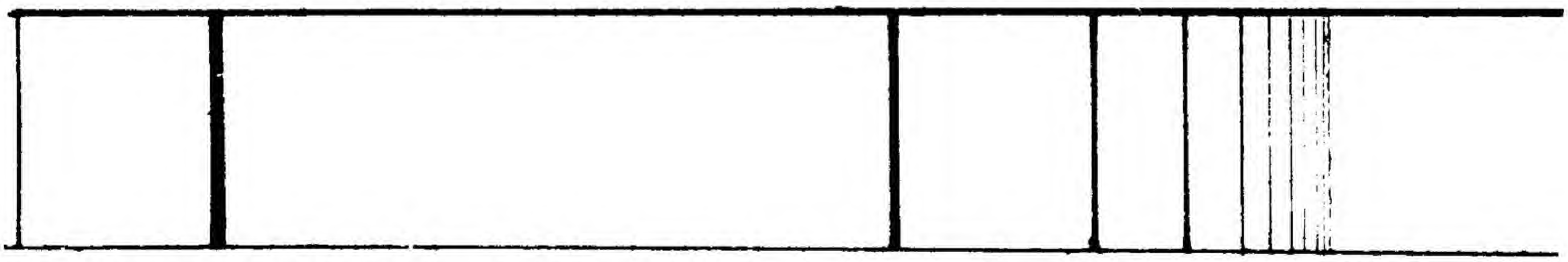


Fig. 3. The Balmer series in the hydrogen spectrum.

(J. W. Mitchell)

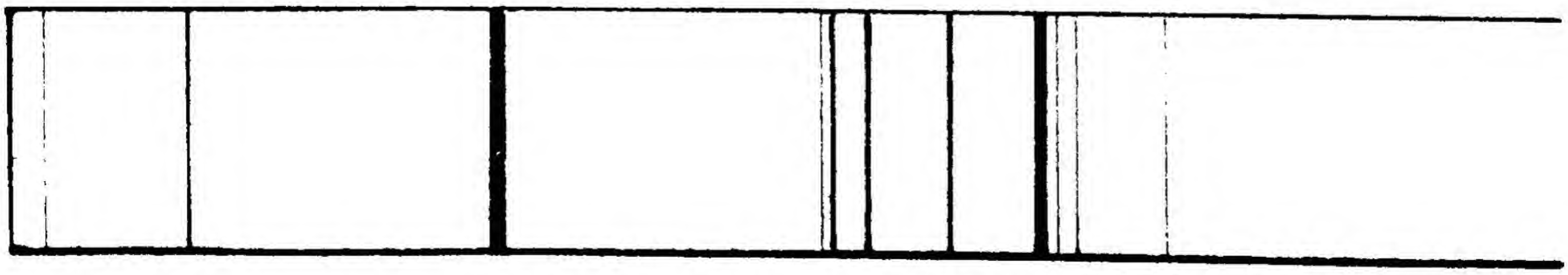


Fig. 4. The helium spectrum.

(J. W. Mitchell)

Fig. 5. The band spectrum of nitrogen

models which will satisfy the above condition. The first regards ordinary light as a transverse wave motion, in which the direction of vibration of any one particle of the medium assumes all possible positions in less than $\frac{1}{20}$ sec. It follows from this that, at any one given instant, the direction of vibration of the different particles of the medium along the ray are different. If the analyser is placed at one point of the ray, the amount of light passing in any $\frac{1}{20}$ sec. will be the same, whatever the direction of the vibration plane passed by the analyser, since the direction of vibration of the particle of the medium at that point assumes all possible positions in that time. Also the eye cannot detect any variations of intensity within a shorter time than that, owing to the persistence of vision. The second, and perhaps more fruitful, view is to regard ordinary light as consisting of two independent plane polarised beams of equal intensity I with vibration planes at right angles to each other. The word, independent, connotes that the two beams are not coherent and so cannot interfere with each other. Therefore the resultant intensity is found *by finding the intensity of each beam and adding the two together*, not by adding the amplitudes algebraically and squaring the result. If the reader asks why the two plane polarised beams are indepenent, the answer is that this condition is necessary to produce ordinary light, which is, after all, what we are trying to do ! If an analyser is placed in a beam of ordinary light with its vibration plane making an angle θ with that of one of the two components of the ordinary light, that component has an intensity I $\cos^2\theta$ on emerging from the analyser, while the other has an intensity I $\sin^2\theta$, and the sum of these two intensities is $I\cos^2\theta + I\sin^2\theta = I$, independent of the angle θ and therefore of the position of the analyser.

151. THE PILE OF PLATES

The pile of plates is another type of polariser and analyser and it is based on the fact that light can be polarised by reflection. It consists of a set of about seven thin glass plates, such as microscope cover slips, placed in contact with their plane at an angle of 33° to the axis of the wooden tube in which they are mounted (Fig. 241), the inside of the tube being blackened to absorb light reflected by the plates.

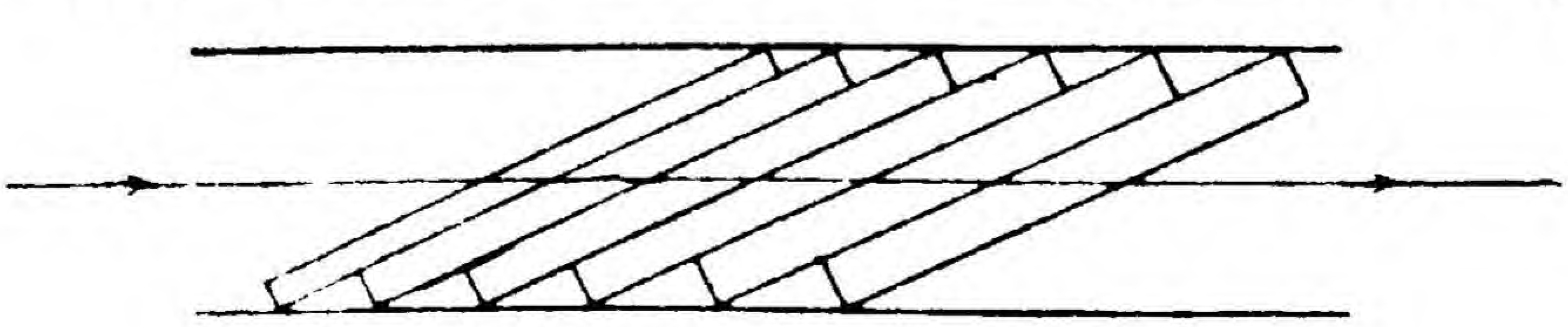


Fig. 241.

If a ray of light parallel to the axis of the tube fails on the pile it will be incident at an angle of 57°, the polarising angle for glass. Hence the light reflected from the first plate will be polarised with the vibration plane at right angles to the plane of the paper. We may regard the incident

ordinary light as two independent components of intensity I with vibration planes in and perpendicular to the plane of the paper. The reflected light is plane polarised, because the glass plate reflects some 10 per cent. of the component whose vibration plane is perpendicular to the plane of the paper, but reflects none of the other component. The transmitted light is therefore partially plane polarised. At the reflection at the second plate, some more of the component with its vibration plane perpendicular to the plane of the paper is reflected and all the other component is transmitted together with still less of the component with vibration plane at right angles to the plane of the paper. The filtering out of this component continues at each successive reflection, until the transmitted light consists entirely of the component whose vibration plane is in the plane of the paper. About seven plates are sufficient to do this and the emergent beam is then plane polarised. Quite good polarisers and analysers can be made in this way and it is now time to use them to investigate and clear up the problem presented by the effect of two calcite crystals on a ray of light.

It will be remembered that, if a ray of light is sent on to a calcite crystal, it is split up into an ordinary ray and an extraordinary ray. If each of these rays is examined with a pile of plates or any other analyser, each is found to be plane polarised, the vibration plane of the ordinary ray being perpendicular to the principal plane of the crystal and that of the extraordinary ray being in that plane. This experiment is merely a deliberate repetition of Malus's observation of the setting sun in a window through a calcite crystal and we see that our guess as to why he saw only one image in certain positions of the crystal is correct. It is clear that the calcite crystal acts as a combination of two slots, one in the principal plane and the other at right angles to it, and that it will only transmit rays whose vibrations are in the directions of these two slots, the one vibrating in the principal plane being transmitted as the extraordinary ray and the other as the ordinary ray. If a second crystal is placed over the first, the amplitude of the ordinary ray must be resolved into two components, one parallel to the principal plane, which will be transmitted as the extraordinary ray, and the other perpendicular to the principal plane, which will go through the crystal as the ordinary ray. If the crystals are orientated identically, the direction of vibration of the ordinary ray from the first crystal is perpendicular to the principal plane of the second and so it goes through the second as an ordinary ray; similarly for the extraordinary ray. Hence only two images are formed, as was seen in Art. 146. When the second crystal is rotated through an angle θ, the ordinary ray of intensity I in the first crystal gives rise to an ordinary ray of intensity $I\cos^2\theta$ in the second producing the image O_{12} in Fig. 237, and an extraordinary ray of intensity $I\sin^2\theta$ giving the image O_1E_2. The extraordinary ray in the first crystal gives rise to an ordinary ray of intensity $I\sin^2\theta$ and an extraordinary ray of intensity $I\cos^2\theta$ in the

second crystal, these rays producing the images E_1O_2 and E_{12} respectively. This general case serves as the basis of the explanation of any other possibility which may occur with two crystals. How simple do these complicated facts appear, when we have found the clue which is needed to unravel them, and how amazing it is that Newton should have been able to see so clearly the essential nature of the underlying conceptions! His two statements have now been proved correct, namely, that the facts demonstrate the one-sidedness of light and that the ordinary and extraordinary rays only differ from each other by some property which can be expressed as a right angle.

152. DOUBLE REFRACTION

It is now time to turn to the other aspect of double refraction, which is the peculiar behaviour of the extraordinary ray. What are the facts? The ordinary ray obeys Snell's law, but the extraordinary ray has a value of $\frac{\sin i}{\sin r}$ which varies with the direction of the ray relative to the optic axis. It is the same as the value for the ordinary ray along the optic axis, and differs most from it in directions at right angles to this axis. There is no double refraction for rays travelling along the optic axis, but two refracted rays or waves are produced in all other cases. On the wave theory, the refractive index of a medium is the ratio of the velocity of light in air to that in the medium, and the fact that the refractive index of the extraordinary ray varies with direction proves that the velocity of the extraordinary wave is different in different directions. Thus, while the secondary wavelets producing the ordinary wave front are spherical, those giving rise to the extraordinary wave front must be of a different shape.

There are two classes of doubly refracting crystal, the uniaxial crystal, such as quartz, tourmaline, and calcite, in which there is only one optic axis, only one direction in which no double refraction occurs, and the biaxial crystal, such as borax, mica, and selenite, in which there are two directions in which there is no double refraction. In future only uniaxial crystals will be considered and Huygens decided to try the following extension of his principle to explain double refraction in such crystals. Each point of the existing wave front is to be considered as the source of secondary wavelets, the wavelets being spheres in the case of the ordinary wave front and ellipsoids of revolution about the optic axis for the extraordinary wave front. The two possible cases are illustrated in Fig. 242; it will be seen that the ordinary wavelet lies wholly within the extraordinary wavelet in the case of calcite, a **negative** crystal, while the opposite is the case for quartz, a **positive** crystal. Calcite is called a negative crystal because, as we shall see, the extraordinary ray is refracted further from the optic axis than the ordinary ray; it is, as it were, repelled by the optic axis. It is seen that the two wavelets touch at the optic

axis, this being necessary to account for the absence of double refraction along this direction. The reader should keep quite clearly in his mind what Huygens was attempting at this stage; he was trying to classify the facts about the extraordinary ray, to make them follow from the

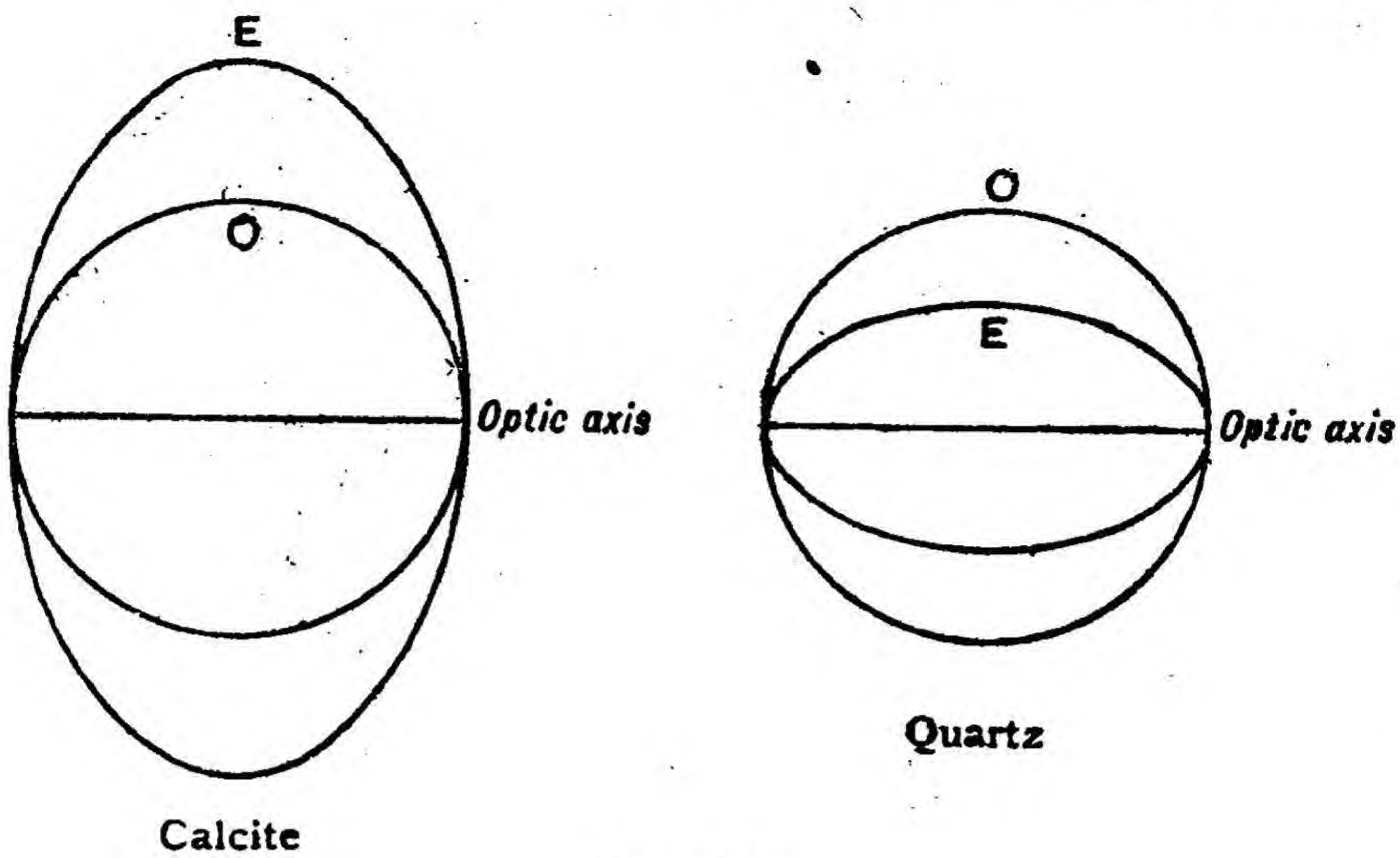


Fig. 242.

above simple extension of his principle. This was not a theory, but a classification of the facts. It remains to be seen if his guess will fit the facts. If it does, it still remains to be seen if we can explain *why* the crystal produces two wavelets and *why* the extraordinary wavelet should be. an ellipsoid of revolution. We shall now consider some cases of double refraction in terms of Huygens' extended principle.

153. HUYGEN'S TREATMENT OF DOUBLE REFRACTION

We will consider first of all the case of a plane wave front incident obliquely on a crystal such as calcite, which has been cut so that the optic axis OA is in the plane of incidence but not parallel to the refracting surface (Fig. 243). The construction follows the usual lines, AB being the incident wave front and CD_o and CD_e being the ordinary and extraordinary refracted wave fronts respectively. The extraordinary ray AD_e is further away from the optic axis than the ordinary ray AD_o, but it does lie in the same plane as the incident ray and the normal to the refracting surface, thus obeying the first part of the law of refraction. But, if the optic axis is not in the plane of incidence, the exercise of a little imagination will enable the reader to visualise the corresponding diagram in three dimensions, which shows that the extraordinary ray is not even in the plane of incidence and so its behaviour is abnormal in every respect.

Fig. 244 shows the case of a plane wave falling obliquely on a surface cut so that the optic axis is in the plane of incidence and parallel to the refracting surface. The results are similar to the previous case and it may

be mentioned that the extraordinary ray AD_e is not even normal to the extraordinary wave front CD_e, as was also the case in the previous instance.

The next case is that of a plane wave falling obliquely on a surface cut

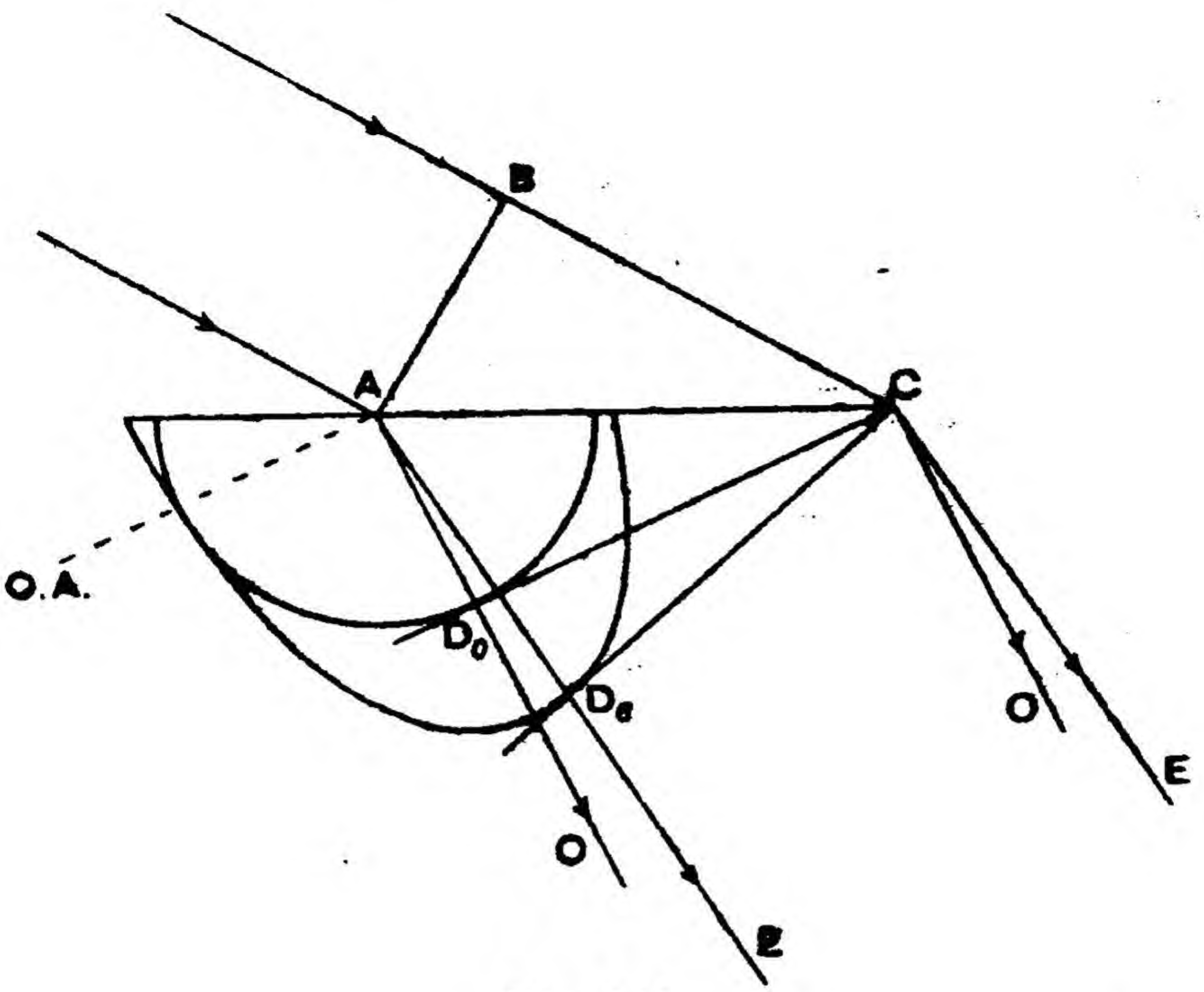


Fig.-243.

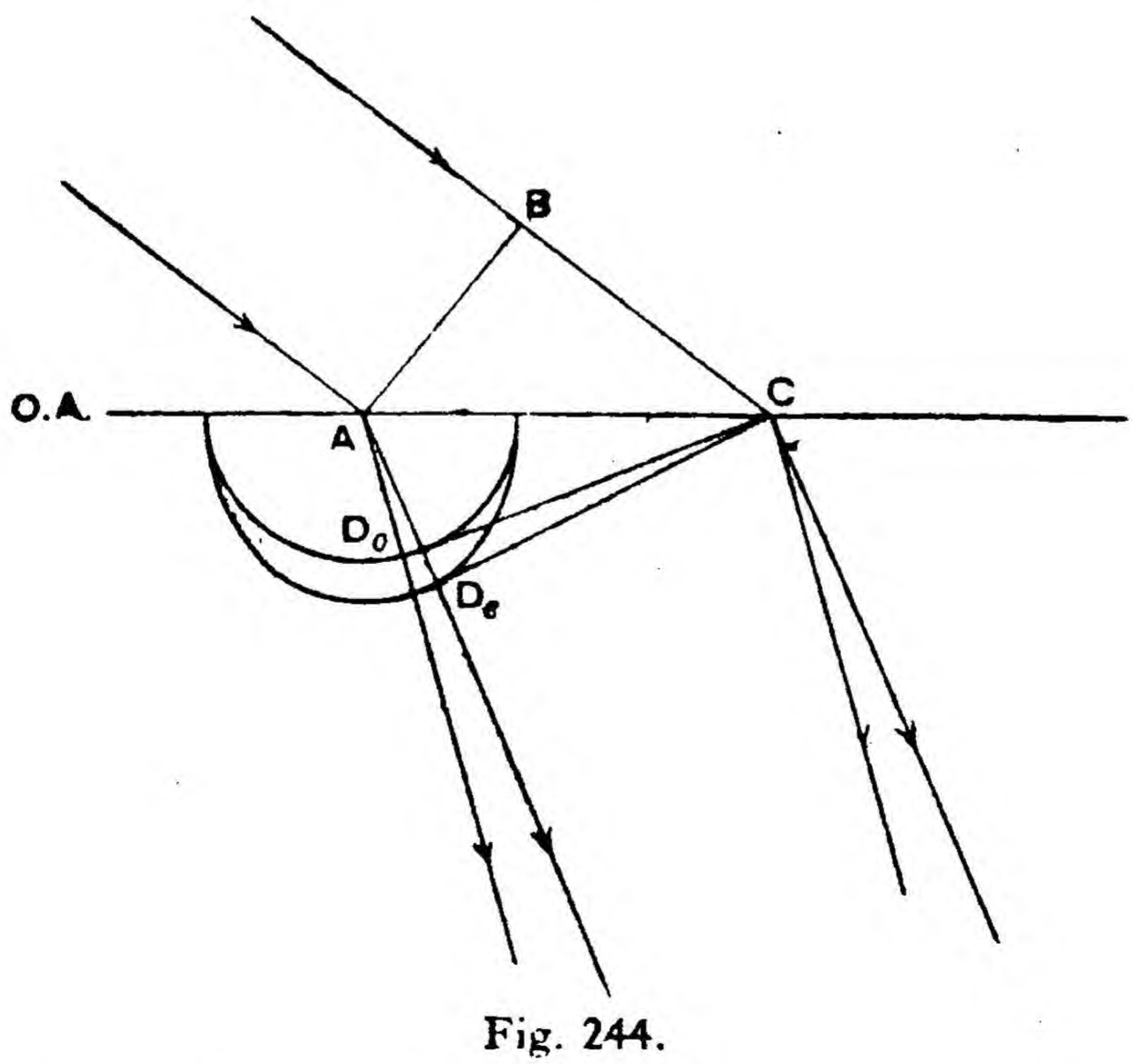


Fig. 244.

so that the optic axis is normal to the plane of incidence (Fig. 245). Since the extraordinary wavelet is an ellipsoid of revolution about the optic axis as axis, the section of this ellipsoid by the plane of incidence is a circle. Hence the sections of both the ordinary and extraordinary wavelets in the

plane of incidence are circles and so each ray obeys Snell's law. This gives us a way of defining the **extraordinary refractive index** of a material as the ratio of the sine of the angle of incidence to that of the angle of refraction in the special case when the incident ray lies in a plane

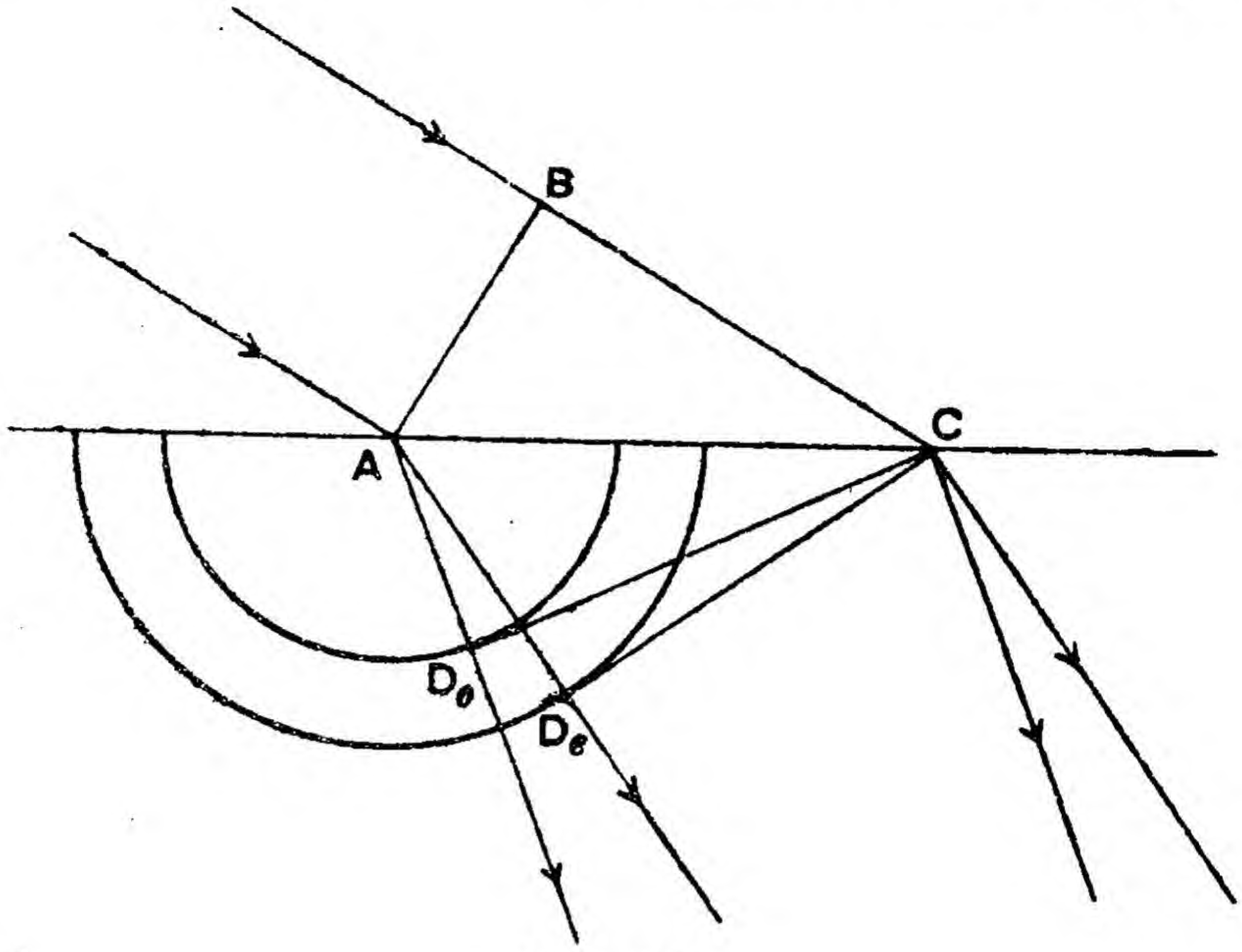


Fig. 245.

perpendicular to the optic axis. The refractive indices of a number of doubly refracting substances are given in Table 16.

TABLE 16

Substance.	n_o	n_e
Calcite.	1·658	1·486
Quartz.	1·544	1·553
Tourmaline.	1·64	1·62
Ice.	1·306	1·307

The case of rays falling normally on a surface with the optic axis in the plane of incidence but oblique to the surface is illustrated in Fig. 246 and shows that the ordinary and extraordinary wavelets may be parallel and yet the two rays are separated due to the extraordinary ray not being normal to its wave front. Rays incident normally on a surface with the optic axis in the plane of incidence and parallel to the surface are shown in Fig. 247 and here both the wave fronts and rays are parallel to each other. Although there is no separation of the rays in direction, there *is* double refraction, for the rays travel at different speeds along the same line and so the wave fronts do separate. This is an important case and the reader must be sure that he understands it. He should draw for himself the case of normal incidence on a surface which is at right angles

to the optic axis, when he will obtain a result similar to that just mentioned. Finally the case of rays incident normally on a surface with the

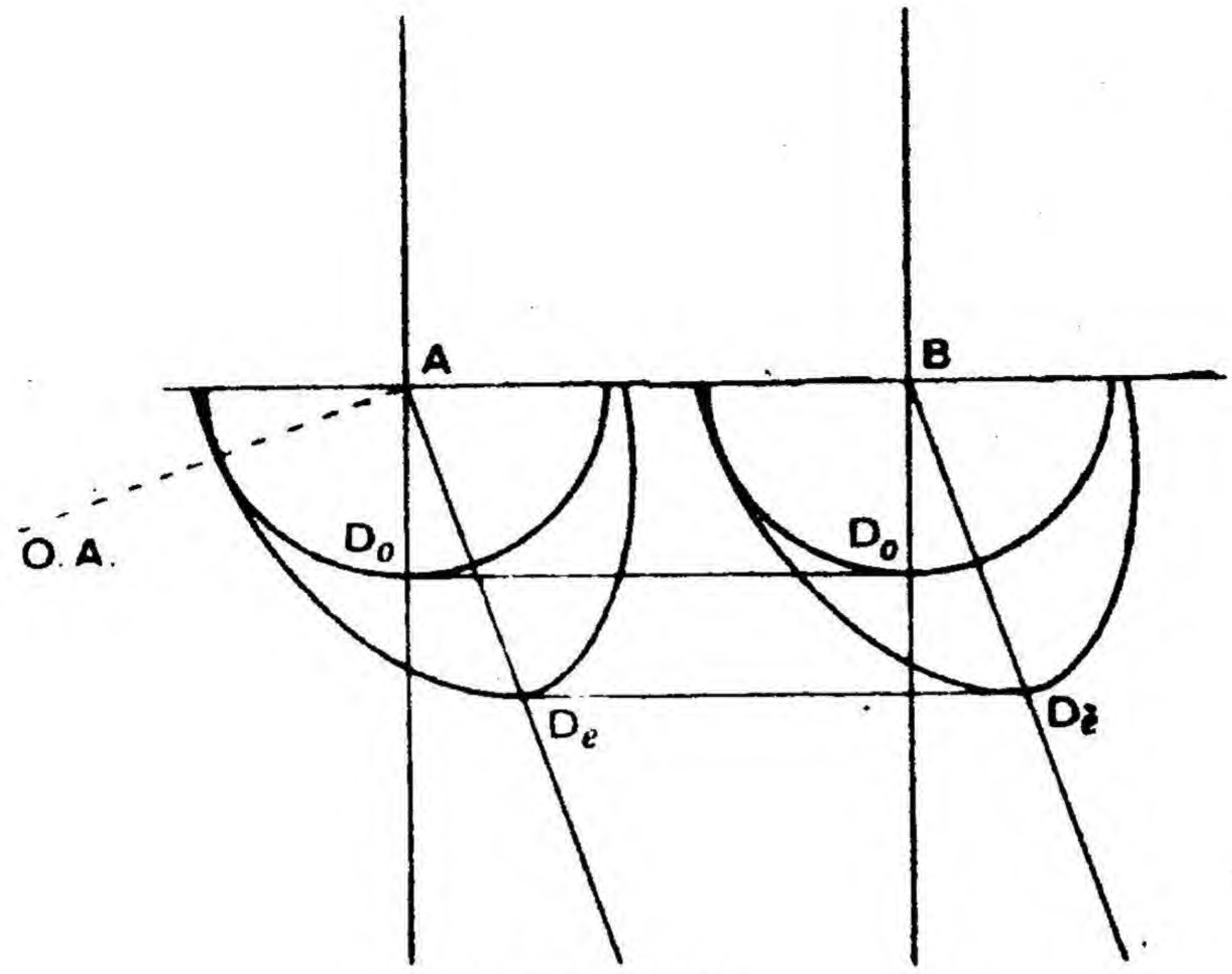


Fig. 246.

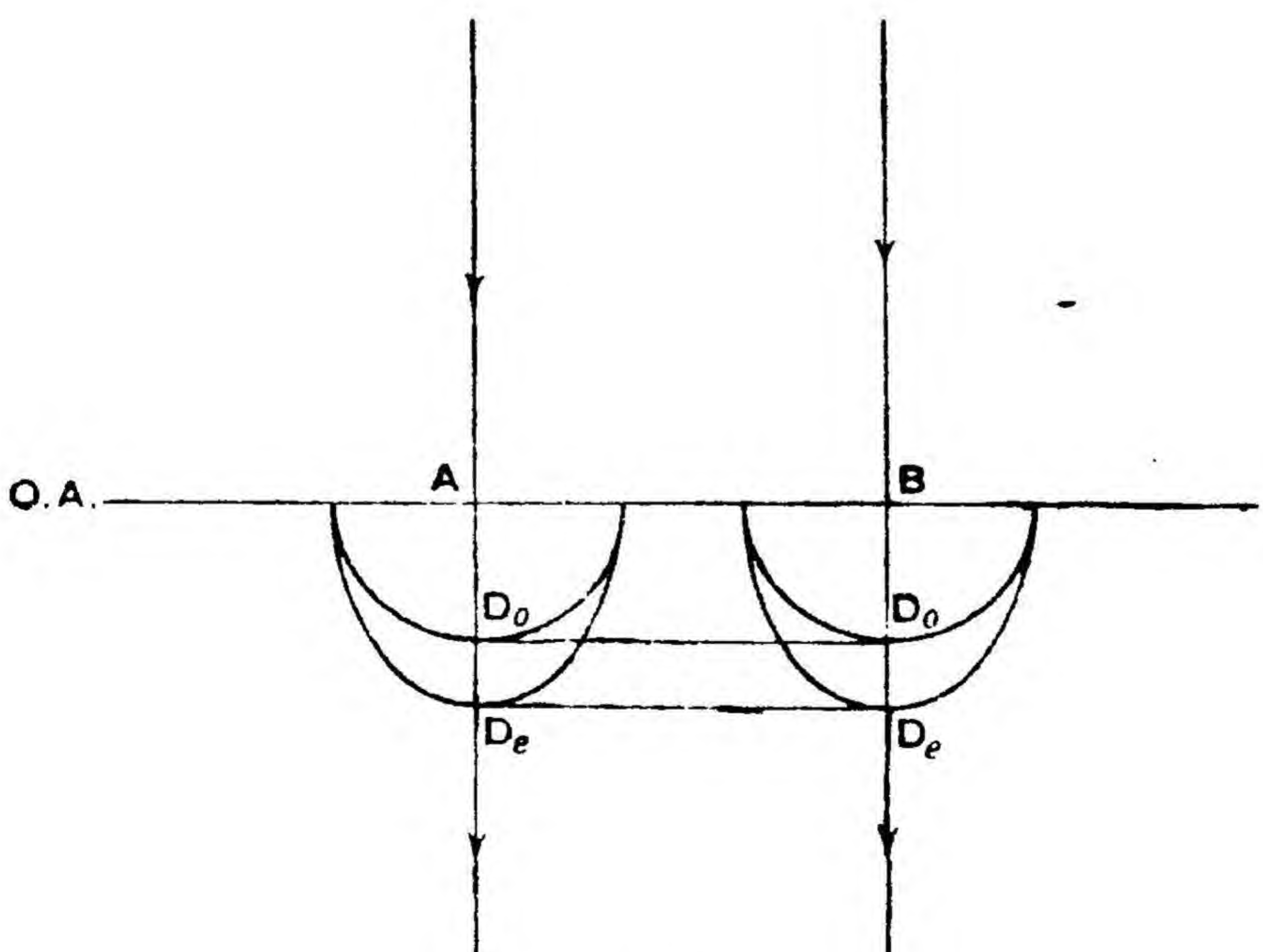


Fig. 247.

optic axis also normal to the surface is shown in Fig. 248 and it will be seen that there is no double refraction in this case ; neither the rays nor the wave fronts separate. It has been stated already that this absence of double refraction along the optic axis is an experimental fact and it merely remains to add that the remainder of the above predictions from Huygens' extended principle are likewise verified experimentally. But Huygens was not able to take the next step of explaining why the crystal

produces two wavelets and why the extraordinary wavelet is an ellipsoid of revolution. This explanation is beyond the scope of this book and we shall not go into it.

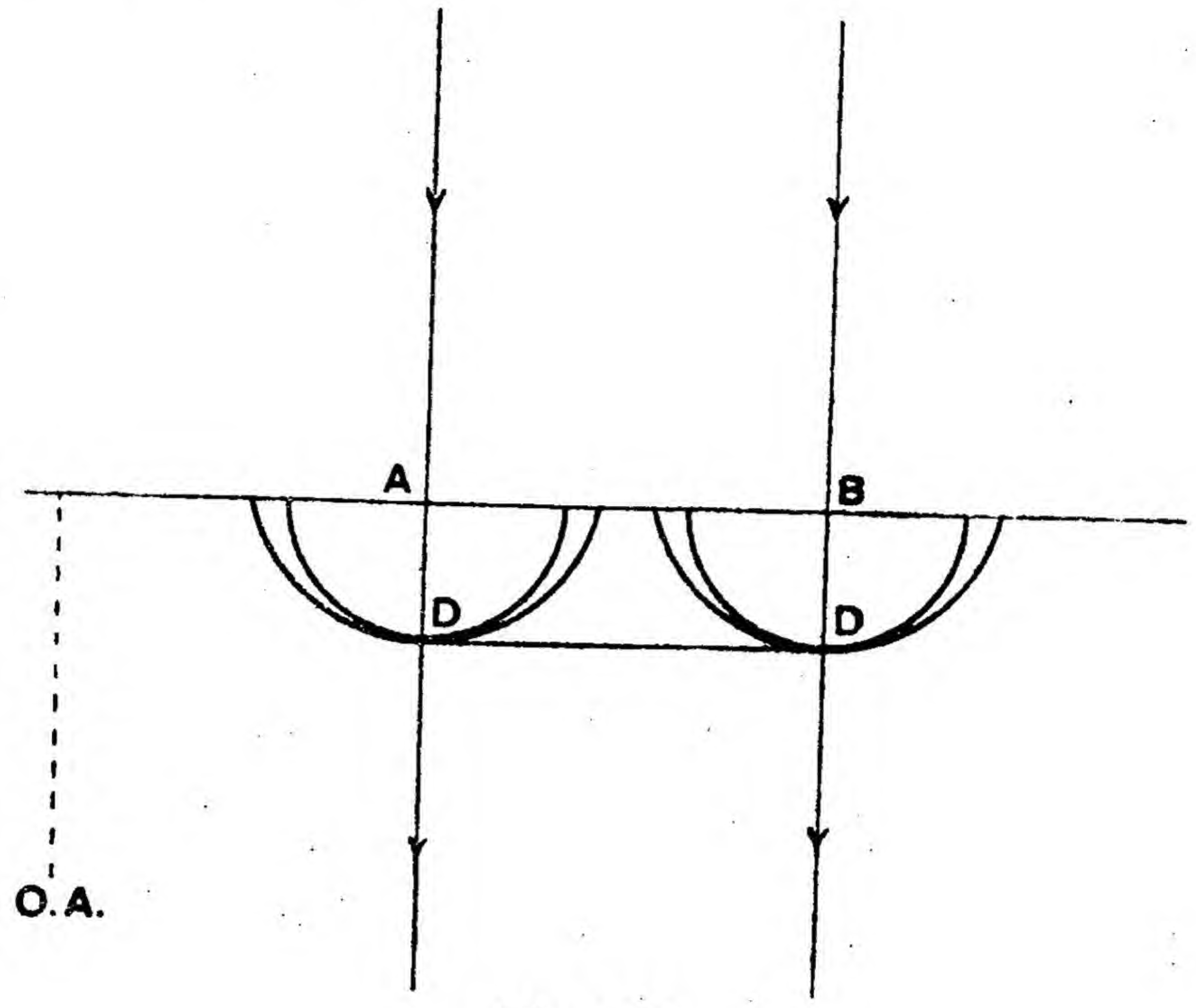


Fig. 248.

154. THE NICOL PRISM

Any device for producing plane polarised light from ordinary light must reduce the intensity to half of its initial value, since it eliminates one of the two independent plane polarised components of which ordinary light may be assumed to consist. But the two polarisers which we have described reduce the intensity still more than this. The tourmaline plate absorbs the extraordinary ray to some extent as well as the ordinary ray completely, while the pile of plates absorbs a considerable proportion of that component which finally emerges plane polarised. So it is desirable to design some device which will produce plane polarised light with the minimum of absorption and two such devices have been produced. One is about a hundred years old, while the other has only been invented in the last three years. The first is the Nicol Prism. It is made by cutting a crystal of calcite (Fig. 234) into two halves by a plane along the line AG, joining the two corners containing three obtuse angles, and perpendicular to the principal section ABGH of the crystal. These two portions of the crystal are then cemented together again with Canada balsam, when the whole looks as transparent as it did before it was cut. But the refractive index of Canada balsam, 1·55, is intermediate between that of calcite for the ordinary and extraordinary rays (Table 16). If a ray of light falls on the crystal as shown in Fig. 249, it is split up into the ordinary and extraordinary rays, the latter making the greater angle with the optic axis. The ordinary ray now falls on the Canada

balsam, an optically less dense medium, at an angle greater than the critical angle, and is totally reflected and goes to the side of the crystal, where it is absorbed by the blackened tube in which the crystal is mounted.

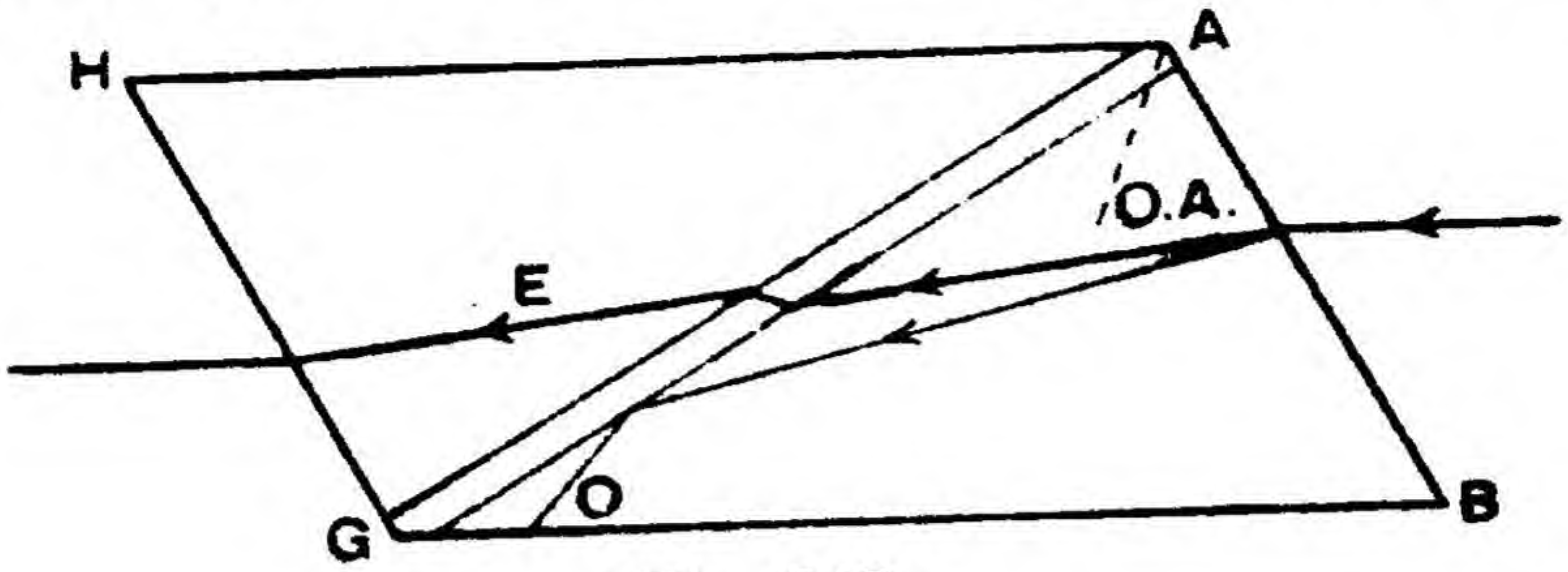


Fig. 249.

But the extraordinary ray is refracted into the Canada balsam, since it is an optically less dense medium for this ray. It cannot be totally reflected on reaching the other half of the calcite, since the original angle of incidence in the first portion of the calcite is less than 90°. Therefore the angle of refraction in the Canada balsam at each refracting surface is less than the critical angle, and the extraordinary ray goes right through the crystal and emerges as a plane polarised ray with its vibration plane in the principal plane of the crystal. If we look at a crystal end on, this plane is parallel to the diagonal AF (Fig. 234). Such a crystal, mounted in a blackened tube to absorb the ordinary ray reflected to one side, is called a Nicol Prism and it is one of the most efficient polarisers known. Its only drawback is its expense, since calcite crystals large enough to pass a beam of light an inch wide are rare and narrow beams must be used. Therefore bright images cannot be obtained if a Nicol Prism is used in the formation of the image. There is one further small point to be noticed. If the incident ray makes a much smaller angle with BA than that shown in Fig. 249, the ordinary ray will fall on the Canada balsam at less than the critical angle and will be transmitted, while, if the incident ray makes a much larger angle with BA than that shown, the extraordinary ray will be so nearly parallel to the optic axis that its refractive index will not be greatly different from that of the ordinary ray. Hence the Canada balsam will be an optically less dense medium even for the extraordinary ray and it will be totally reflected. Therefore parallel rays having a direction about that shown in Fig. 249 should be sent through a Nicol prism if completely plane polarised light of the maximum intensity is required.

155. POLAROID AND ITS POSSIBILITIES

The more recent polariser referred to in the previous article is called polaroid. It is effectively a large single crystal of herapathite, which is a periodide of quinine sulphate. The crystals of this substance are doubly refracting and, like tourmaline, absorb the ordinary ray, transmitting only the extraordinary ray. A large single crystal is produced

by suspending ultra-microscopic crystals in nitro-cellulose with a layer of pure nitro-cellulose above and below the suspension. The suspension is then forced under pressure through a fine horizontal slit, which orientates all the tiny crystals in the same direction in the liquid. The tiny crystals act effectively as a single crystal and the suspension is mounted between two glass plates, the vibration plane of the transmitted light being marked on the frame in which the plates are fixed. Discs of polaroid an inch in diameter can be obtained for two or three pounds and discs five or six inches in diameter can be bought. Polaroid promises to be an excellent substitute for all the purposes for which Nicol Prisms are used at present, since it is much cheaper and is just as efficient a polariser except at the extreme ends of the spectrum.

It has an interesting commercial possibility. The dazzle of oncoming motor car headlights is responsible for a number of serious road accidents and a way of overcoming this dazzle is urgently needed. This can be done, if all headlights are fitted with discs of polaroid arranged so that the vibration plane of the light transmitted is at 45° to the vertical in a clockwise direction, as seen by the person driving the car. If the driver of a car A looks at the road through a small piece of polaroid with its vibration plane parallel to those of the discs over his own headlights, he will see any objects illuminated by the light from his own headlights, since the reflection of that light back to him will not change its vibration plane. But the polaroid disc through which he is looking will stop any light from the headlights of an oncoming car B, since the vibration plane of the light from B's headlights will be at 45° to the vertical in an *anti-clockwise* direction as seen by the driver of car A and so dazzle is eliminated. The reason for this is that the vibration plane of the polaroid discs over the headlights of car B was 45° to the vertical in a clockwise direction as seen by the driver of A, when the two cars were side by side facing the same way. When B is turned round so as to approach A, the vibration plane of its polaroid discs is now 45° to the vertical in an *anti-clockwise* direction as seen by A's driver and so it is at right angles to the vibration plane of the disc through which he is looking. This device has been successfully demonstrated under road conditions, but it has not yet been adopted commercially. It may become general when the discs for the headlights can be produced at a reasonable price.

156. ROTATION OF THE PLANE OF POLARISATION

Our investigations have been inspired up to the present by the question: are light waves longitudinal or transverse? We have seen that they are wholly transverse and our investigations have led us to the conception of plane polarised light, its production, and detection. It is not long after a new idea of this kind is established before it is made use of in turn to suggest new ideas and applications. We have already seen how the

conception of interference was investigated in order to see if light was waves and that now light waves themselves are being used to measure the thickness of thin films and even as a standard of length. It was found by Arago in 1811 that, if a ray of ordinary light is passed through a plate of quartz cut normal to the optic axis between a polariser and analyser with vibration planes at right angles to each other, the emergent beam is not extinguished as is the case in the absence of the quartz. A polariser and analyser with vibration planes at right angles to each other are said to be crossed, and so the phrase **crossed Nicols** means Nicols with their vibration planes perpendicular to each other and passing no light in the absence of any other doubly refracting material in between them. Arago showed that the emergent light could be extinguished once more by a mere rotation of the analyser through a definite angle. Therefore quartz rotates the vibration plane of polarised light which passes through it along the optic axis. This rotation of the vibration plane is found to be proportional to the thickness of the quartz and inversely proportional to the square of the wave-length of the light. A plate of quartz 1 mm. thick at 20° C. produces a rotation of 16·4° with red light, 21·7° with yellow light, and 47·5° with violet light. The rotation increases with temperature. Some specimens of quartz rotate the vibration plane in a clockwise direction when seen by an observer looking at the quartz towards the source of light, and are called **dextro-rotatory**, while others produce an anti-clockwise rotation as seen by an observer similarly placed and are called **lævo-rotatory**.

This property of rotating the vibration plane of polarised light is also exhibited by certain liquids and solutions, such as turpentine, solutions of sugar and tartaric acid. The rotation of a given wave-length by a solution at a given temperature is proportional to the length of the solution and its concentration in grams per c.c.. The rotation θ due to a solution l cm. long of concentration c gm. per c.c. is given by

$$\theta = \frac{s \times l \times c}{10}$$

where s is the **specific rotation** of the solution at the given temperature for the given wave-length. It is the rotation produced by a solution 10 cm. long of concentration 1 gm. per c.c. Some solutions are dextro-rotatory and others lævo-rotatory, and, as with quartz, different varieties of the same solute can give rise to dextro-rotatory and lævo-rotatory solutions. It is evident that this phenomenon can be used to find the strength of a solution, if its specific rotation has been measured. All that is needed for this purpose is a pair of crossed Nicols or a similar polarising device ; the analyser is set in the position to extinguish the emergent beam and the solution is inserted. The analyser is then re-set so as to extinguish the emergent beam once more and, if it is provided with a circular scale marked in degrees, the rotation of the solution of known length can be

measured and its strength can be calculated. If this method is tried in practice using just crossed Nicols, it will be found that it is impossible to set the analyser to more than one or two degrees, so there will be an uncertainty of two to four degrees in the value of the rotation. This is too large for practical purposes if the rotation is about 40°, so some device is needed which will enable the analyser to be set to a quarter of a degree, since the change in intensity produced by this amount of rotation with crossed Nicols is too small to be detected. One such device will be described; there are a number of polarimeters, as they are called, but the reader must consult more advanced books if he wishes to find details of them. All of them aim at the same thing, an accurate setting of the analyser in the crossed position.

157. THE CORNU-JELLET HALF-SHADE POLARIMETER

This consists of an ordinary Nicol Prism P (Fig. 250) as the polariser, a tube T to hold the solution under test, and a special analyser A. This is

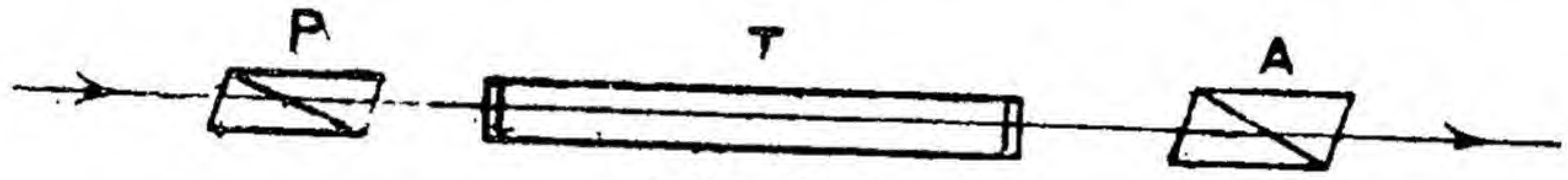


Fig. 250.

a Nicol Prism, ABFE, the letters referring to the same points of the calcite crystal illustrated in Fig. 234. This prism is shown end-on in Fig. 251; a wedge-shaped piece, represented by PQR, is cut out of the prism and the two halves labelled 1 and 2 are cemented together again, so that the vibration planes of the light passed by each half make a small angle with each other. If a ray of monochromatic light is sent through the polarimeter without any solution in the tube the analyser can be rotated to such a position that each half of the field is equally bright. To understand how this can be done, consider Fig. 252, in which XOX′ is the direction of vibration of the light passed by the polariser and OV_1 and OV_2 are the directions of vibration passed by the half 1 and half 2 respectively of the analyser. If the analyser is turned so that OV_1 is normal to XOX′, then half 1 is dark and half 2 is faintly illuminated. Conversely the small clockwise rotation needed to make OV_2 perpendicular to XOX′ transforms the field of view so that half 1 is now faintly illuminated while half 2 is dark. This marked transformation occurs for a rotation of two or three degrees and half-way in between is a position so that the angle between OV_1 and OV_2 is bisected by YOY′, the line perpendicular to XOX′. In this case each half of the field is

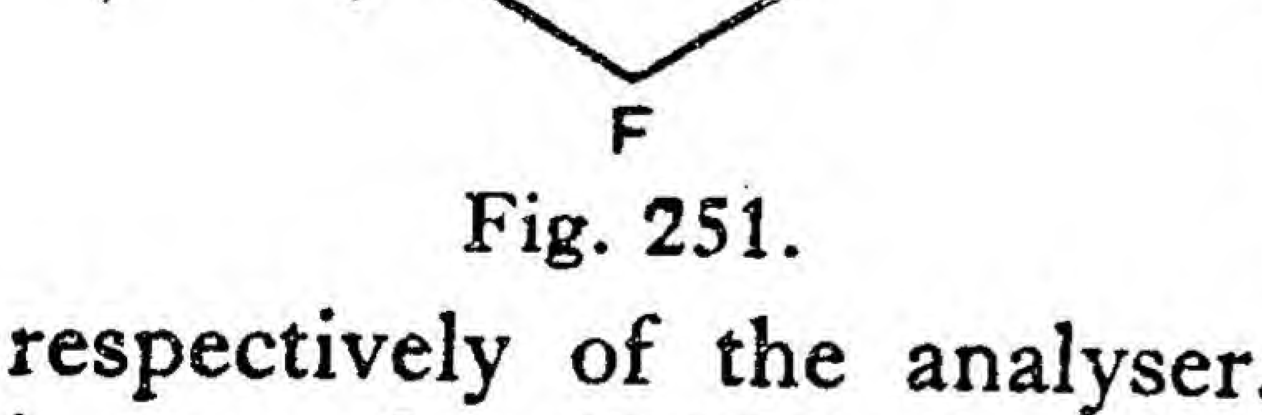


Fig. 251.

equally bright. This setting can be made to a quarter of a degree and the position of the analyser is read on a circular scale marked in degrees, which is not shown in Fig. 250. The tube of solution, whose rotation is to be measured, is now placed in position and the vibration plane of the light emerging from the tube will no longer make equal angles with OV_1 and OV_2 and the two halves of the field will look unequally bright. The analyser is then set to make each half equally bright once more and its new position is read. The difference between these two readings gives the angle through which the tube of solution has rotated the vibration plane and from which the strength of the solution can be calculated, if its specific rotation is known.

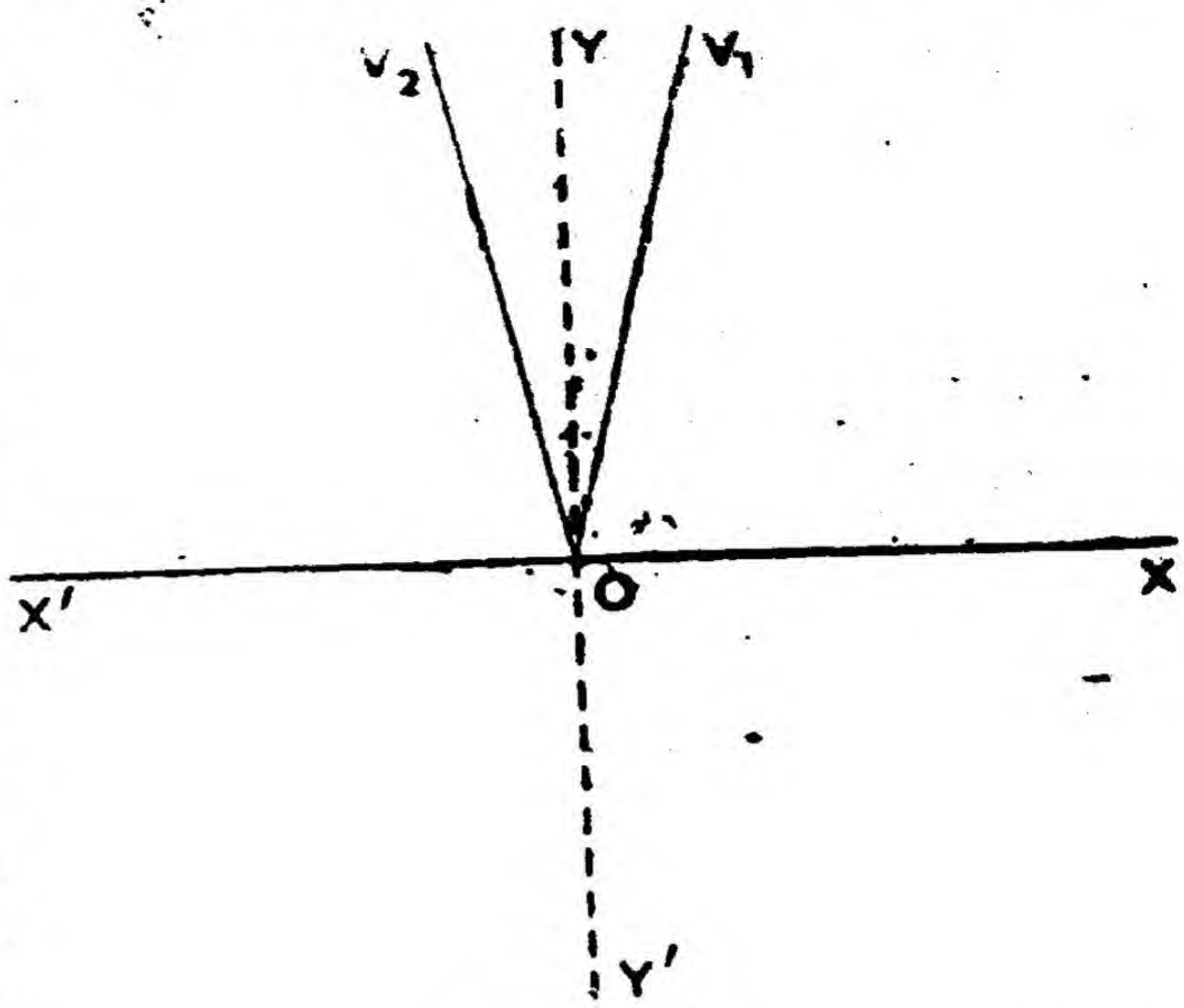


Fig. 252.

158. FURTHER POLARISATION PHENOMENA

Faraday was always convinced that there was a connection between magnetism and light and he actually sought for the Zeemann effect, which concerns the way in which spectral lines are modified, when the atoms emitting them are in a strong magnetic field. His apparatus was too crude to detect the effect, which was discovered by Zeemann some sixty years later. But he did discover one connection between magnetism and light by showing that a magnetic field can cause rotation of the vibration plane of polarised light. If a plate of glass is placed between the poles of an electro-magnet, whose pole pieces are bored so that light can pass through the glass in the same direction as the lines of force, and crossed Nicols are placed on either side of the glass, the light is extinguished in the absence of a magnetic field. As soon as the electro-magnet is excited, the light appears again, but it can be extinguished by a suitable rotation of the analyser. The angle α through which the vibration plane is rotated is proportional both to the length l of the field and its strength H, so we have

$$\alpha = clH$$

where c is called Verdet's constant. It is the rotation produced by a field of 1 gauss 1 cm. long. It is $3{\cdot}4''$ for lead glass, $4{\cdot}4''$ for molten sulphur, $2{\cdot}5''$ for carbon disulphide. The vibration plane is usually rotated in the same direction as the flow of the electric current, which would produce the magnetic lines of force causing the rotation. Some substances produce a so-called negative rotation in the opposite direction to this. It follows that the direction of the rotation for a given observer is

the same whichever way the light travels, and, if a ray of light is sent to and fro along the same lines of force, the rotation is increased at each traverse of the field and the total rotation is greatly magnified by this repeated reflection. The opposite is the case with optically active substances, such as sugar solutions, where the direction of rotation as seen by a given observer is reversed when the ray is reversed, so that the rotation is annulled if a ray comes from a source to the observer and is reflected back through the solution towards the source again.

The **Kerr effect** was discovered by John Kerr in 1865. Normally glass is isotropic, but, if a piece of glass is placed between the plates of a condenser and an electric field is switched on it becomes doubly refracting. If a ray of light is sent through the glass, it is split up into two components travelling at different velocities, one polarised with the vibration plane along the electric lines of force and the other with the vibration plane at right angles to them. The path difference n in wave-lengths between the two components after passing through an electric field of strength E volts per cm. of length l is given by

$$n = klE^2$$

where k is the Kerr constant for the substance used. Its value for nitro-benzene, a common substance exhibiting a strong Kerr effect, is $2{\cdot}7 \times 10^{-8}$. A condenser with nitro-benzene between the plates is called a **Kerr cell** and, if such a cell is placed between crossed Nicols, no light will emerge in the absence of an electric field. If a field is switched on between the plates of the condenser, the nitro-benzene becomes doubly refracting and some light emerges from the analyser. Let us imagine that the vibration plane of the polariser is at 45° to the lines of force of the field, which are along XOX′ (Fig. 253). The light incident on the Kerr cell represented by OB is resolved into two components of equal amplitude, one vibrating parallel to the lines of force represented by OC, the other vibrating perpendicular to the lines of force represented by OD, travelling at different speeds. When they emerge from the cell, there will be a phase difference between them; let it be π, which would be the case if one component had gained half a wave-length on the other. The two vibrations are now represented by OE and OD. If the analyser is crossed with the polariser, the component of each of these vibrations in

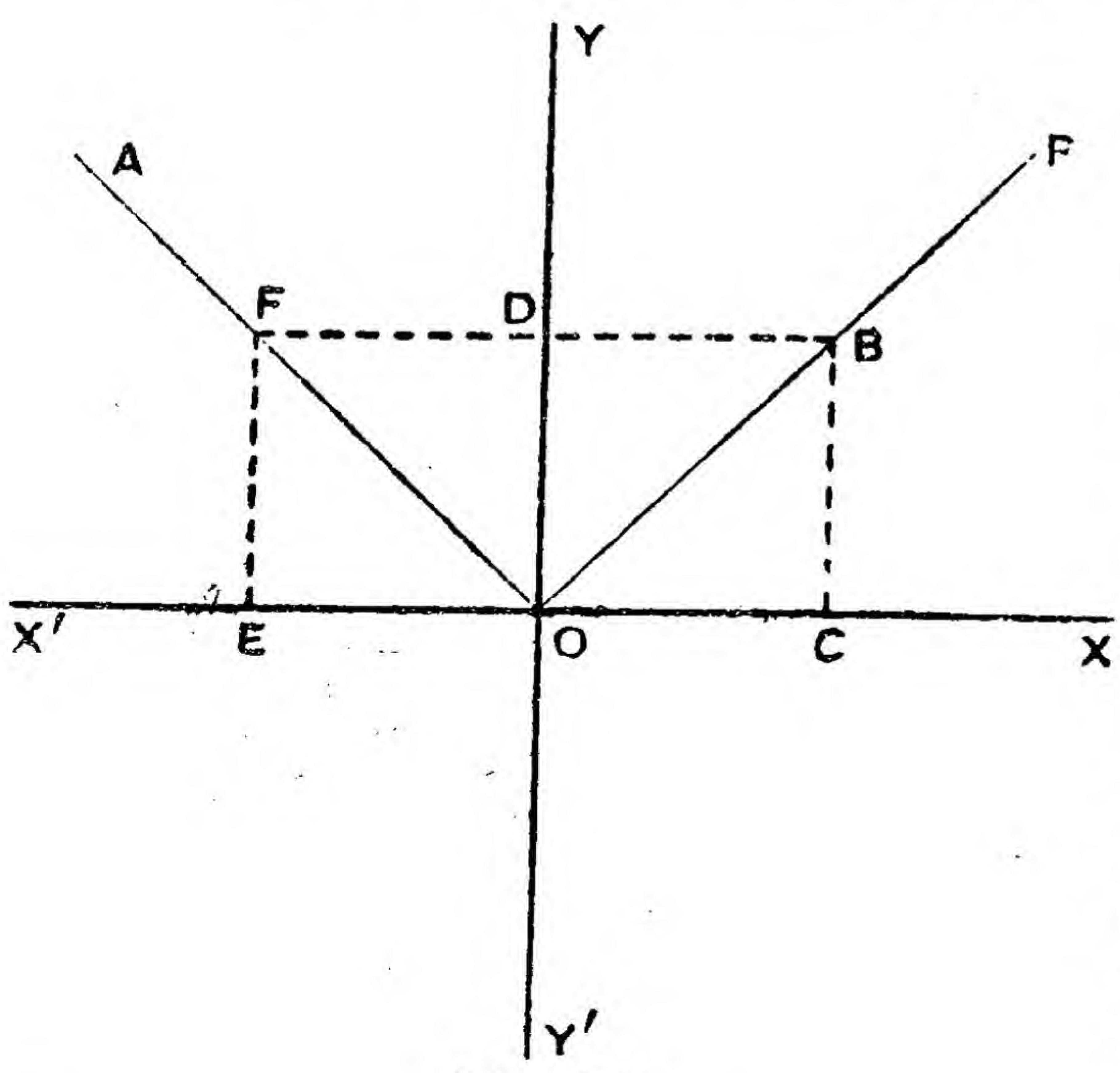


Fig. 253.

the direction OA of the analyser will be in phase and so light of finite amplitude, represented by OF, will emerge from it. This power of the Kerr cell to pass light through crossed Nicols by applying an electric field to the cell can be used to change varying electric currents into beams of light of varying intensity and it has been employed both in the transmission of pictures by telegraph and in television.

Finally, polarised light is being used in engineering to detect the presence of strains in structures such as bridges. If an unstrained piece of glass is placed between crossed Nicols, no light comes out of the Nicols, but if the glass is strained it becomes doubly refracting and some light emerges from the analyser. The presence of strain at a particular place is revealed by the presence of light at that place in the glass and the extent of the strain can be estimated from the colour, if white light is used. It is important that the glass used for making lenses should be free from strain and it is tested by examination between crossed Nicols. The presence of strain in a structure, such as a bridge, is tested by building a model of the bridge in an easily strained transparent material, such as zylonite, and examining the model between crossed Nicols. It is possible to deduce the magnitude of the strain at each part of the model from the appearance of the field and therefore the strain to be expected at the corresponding parts of the actual structure. The technique of this important advance in engineering was initiated by Professor Coker, who is an authority on photo-elasticity. The strains to be expected in a bridge of reinforced concrete to be placed over the Rhône at Balme were investigated in this way by Mesnager.

159. CIRCULARLY AND ELLIPTICALLY POLARISED LIGHT

We have already seen that the ordinary and extraordinary rays travel along the same paths at different speeds, when a ray of light falls normally on a plate of doubly refracting material cut so that the optic axis lies in the refracting surface (Art. 153, Fig. 247). Since the waves travel at different speeds, they become separated and a phase difference of increasing value is produced between the ordinary and extraordinary waves at the same point of the crystal, as the wave travels on through it. If the crystal is so thick that the phase difference is $\frac{\pi}{2}$ on emergence, it is called a **quarter-wave plate**, because one wave gains a quarter of a wave-length over the other in passing through the crystal. If light falls on crossed Nicols, with a quarter-wave plate placed between them so that the vibration planes of the ordinary and extraordinary rays in it are at 45° to those of the polariser and analyser, the intensity of the light emerging from the analyser remains the same if the analyser is rotated. The quarter-wave plate seems to have converted the plane polarised light into ordinary light once more! It seems to have depolarised the plane

polarised light! Fresnel first discovered this when investigating the effect of total reflection at a glass-air surface on polarised light.

What is the nature of this depolarised light? Since the vibration plane of the light incident on the quarter-wave plate is at 45° to the two vibration planes passed by the quarter-wave plate, the incident light is resolved into two vibrations of equal amplitude and zero phase difference in two perpendicular directions. These vibrations emerge from the quarter-wave plate with a phase difference of $\frac{\pi}{2}$ and we have already seen (Art. 120) that two vibrations of the same amplitude and period with a phase difference of $\frac{\pi}{2}$ along lines at right angles to each other combine into a circular vibration. Therefore the light emerging from the quarter-wave plate is called circularly polarised light, each particle of the medium, along which the light ray passes, describing a circle of the same radius in the same time. The analyser naturally produces no variation in intensity of the light, since such a circular motion can be resolved into two simple harmonic motions in any two mutually perpendicular directions. How can this circularly polarised light be distinguished from ordinary light? It is done by passing each through a quarter-wave plate. The ordinary light will be unaffected, since it may be regarded as two independent components plane polarised in the two vibration planes of the quarter-wave plate and the phase difference introduced is immaterial, since the two components are incoherent. But the circularly polarised light will be resolved into two simple harmonic vibrations along the two vibration planes of the quarter-wave plate and a further phase difference of $\frac{\pi}{2}$ will be introduced, so that the two components will emerge with a total phase difference of π and will therefore combine into a single simple harmonic vibration (Art. 120). Therefore the passage through a quarter-wave plate does not affect ordinary light, but it converts circularly polarised light into plane polarised light, which can be distinguished from ordinary light in the usual way. It is interesting that the quarter-wave plate depolarises plane polarised light and can also polarise this "depolarised" light!

Since two simple harmonic motions of unequal amplitude and the same period executed in perpendicular directions with a phase difference of $\frac{\pi}{2}$ combine to an elliptical motion, elliptically polarised light can be produced by putting a quarter-wave plate between crossed Nicols with its vibration planes at any angle to those of the polariser and analyser other than 45°. The light emerging from the wave plate will be elliptically polarised and, if it is examined in the analyser, it will give two maxima and two minima not of zero intensity in each revolution. It thus resembles

partially plane polarised light. It can be distinguished from partially plane polarised light by passing it through another quarter-wave plate with its vibration planes parallel to those of the first one, when the two linear vibrations, into which the elliptical motion is resolved, will have their phase difference increased to π and will combine to form plane polarised light on emerging from the second quarter-wave plate. But partially plane polarised light is unaffected by a quarter-wave plate and so the emergent light can be distinguished in the two cases.

160. A DETERMINATION OF THE VELOCITY OF LIGHT BY KAROLUS AND MITTELSTAEDT

We have already seen (Art. 106) that the velocity of light may vary with time and the way to test this is to devise an accurate laboratory method, which can be permanently set up for determinations every six months. Fizeau's toothed wheel method could be adopted for this purpose, if only the toothed wheel could rotate at a million revolutions a second. This is out of the question with a mechanical wheel, but it is now possible to replace it with an " electrical toothed wheel " based on the Kerr effect, which will cut off the light intermittently at a rate corresponding to the above speed of rotation of a mechanical wheel. If a Kerr cell is placed between crossed Nicols, it transmits no light in the absence of the field and passes light when the field is switched on (Art. 158). So the field is like a space and the absence of the field like a tooth. The modern thermionic valve used in wireless receivers and transmitters has made it possible to produce varying electric fields with a frequency of 10^7 cycles per second, and to hold this frequency constant to ± 200 cycles per second. If such a field is applied to a Kerr cell between crossed Nicols, the light is alternately passed and stopped 2×10^7 times per second. A time of $\frac{1}{4 \times 10^7}$ sec. elapses between one passage of the light and the next stoppage and light travels a distance of some 7 metres in this time. So the Kerr cell used in this way could be applied to Fizeau's method with a base-line of 3·5 metres.

Karolus and Mittelstaedt used a rather different compensation method, illustrated in Fig. 254, in which a ray of light passes from a source S through crossed Nicols N_1 and N_2, between which are placed crossed Kerr cells, K_1 and K_2. The lines of force of the electric field in each cell are mutually perpendicular and at 45° to the vibration plane passed by the Nicol N_1. If no electric field is applied to either cell, no light is seen by an eye at E. If a steady electric field of the same value is applied to each cell, no light is seen at E again, because the phase difference produced by the first cell between the two rays, into which the incident light is split, is exactly annulled by that due to the second one. If a valve oscillator is now made to supply to each cell a varying electric

field of the same amplitude, frequency, and phase by connecting the plates of each cell in parallel with one of the condensers of the oscillator, still no light will be seen at E, if the distance between the two cells is zero. This is because the light takes no time to pass from one cell to the other and therefore the strength of the field at the second cell, when the light passes through it, is the same as its value when that portion of the light was passing through the first cell. So the double refraction produced

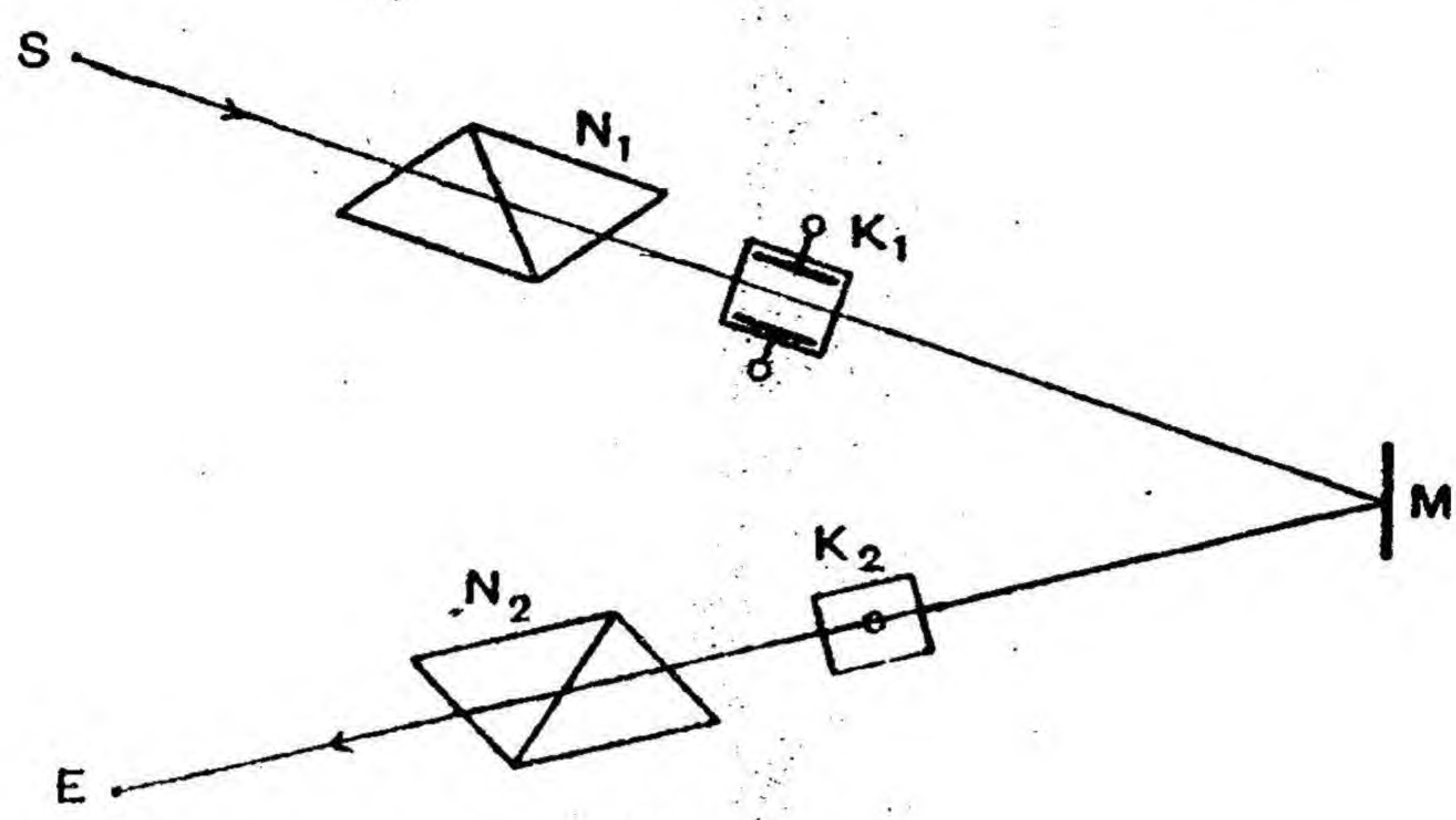


Fig. 254.

by the first cell is compensated out by the second one and no light emerges from the Nicol N_2. If the cells are a finite distance apart, however, some light will emerge from the second Nicol, because the strength of the field at K_2, when a portion of the light passes through K_2, is not the same as its value at K_1, when that same portion of light was passing through K_1. Hence the double refractions of the two cells do not cancel out and some light gets through the system. But darkness can be restored once more by adjusting the frequency of the electric field, so that the light goes from one cell to the other in one cycle of variation of the electric field. If the value of this frequency, f, is found and the distance apart of the cells is d, the velocity of light c is given by

$$c=\frac{d}{\frac{1}{f}}$$

or

$$c=df$$

It is equally possible to adjust the frequency so that k cycles elapse, while the light passes from one cell to the other, when we have

$$c=\frac{d}{\frac{k}{f}}$$

or

$$c=\frac{df}{k}$$

Karolus and Mittelstaedt used a base-line of 300 metres obtained by repeated reflections and a value of *k* from 4 to 8. They obtained the value of 299,778±20 km. per sec. No experiments have yet been done to test the variation of the velocity of light with time, but this method has possibilities and it will doubtless be tried very soon.

EXAMPLES ON CHAPTER XV

1. It has been said that the vibrations of *plane polarised light* take place in a plane perpendicular to the *plane of polarisation*. Say what you understand by the words in italics. On what experimental evidence does this statement depend? (*Oxford Schol.*)

2. Describe how you would produce a parallel beam of plane polarised monochromatic light. Give an account of the principal phenomena exhibited by polarised light. (*Oxford Schol.*)

3. Write a short essay on the polarisation of light. (*Camb. Schol.*)

4. Describe a method of producing a beam of plane polarised light.
How would you distinguish between a beam of plane polarised light and a beam of unpolarised light? Is there any other possibility? (*Camb. Schol.*)

5. Elaborate the statement that, as far as concerns solid transparent media, Snell's law applies only to amorphous substances and to crystals of the highest type of symmetry. (*Camb. Schol.*)

6. A parallel beam of light is incident normally on the plane face of a doubly refracting positive crystal. Construct both the refracted wave fronts and the refracted rays when the optic axis is (a) in the plane of incidence and parallel to the refracting surface, (b) perpendicular to the plane of incidence.

7. Describe with clear diagrams cases of refraction in which (a) the rays are *not* perpendicular to the wave fronts, (b) double refraction occurs but two rays are *not* produced. Would you expect the two rays to show interference on emerging from the crystal? Give reasons for your answer.

8. A horizontal ray of light passes through a Nicol prism with its principal plane vertical and then on to a calcite crystal with its principal plane horizontal. The light then falls on to a screen. What will be seen on the screen and how will the appearance change as the Nicol prism is rotated through 360°?

9. A calcite crystal is placed on a pencil dot on a horizontal piece of paper and a second calcite crystal is placed above the first one. Describe and explain what happens as the upper crystal is rotated about a vertical axis.

10. Describe the manufacture and properties of polaroid and discuss its industrial possibilities.

11. Describe and briefly explain the special feature in the transmission of light through (i) Iceland spar (calcite), (ii) quartz, and (iii) tourmaline.
Explain the construction of a quartz half-wave plate, and calculate the thickness of such a plate for sodium light ($\lambda=5890$ A.U.), given that the appropria[te] refractive indices are 1·5442 and 1·5533. (*London B.Sc.*)

12. A parallel beam of sodium light is incident on a 60° calcite prism cut with its optic axis parallel to the refracting edge of the prism. If the prism is set so that the ordinary ray passes through it at minimum deviation, calculate the distance between the ordinary and the extraordinary image of the D line, if the spectrum is focussed on a screen with a converging lens of focal length 50 cm. Find also the angle through which the prism must be turned so that the extraordinary ray passes through it at minimum deviation. The extraordinary and ordinary refractive indices of calcite for sodium light are 1·486 and 1·658 respectively.

You discover that calcite is specially transparent to infra-red radiation. In view of the above results, how would you cut a calcite prism for producing and measuring infra-red spectra ?

13. What do you understand by the statement that a solution of cane sugar in water is dextro-rotatory to polarised light ? Describe how you would measure the effect in the laboratory. (*Oxford Schol.*)

14. Describe, and give the theory of, a type of polarimeter suitable for measuring accurately the rotation produced by optically active solutions, and explain why a simple arrangement of polarising and analysing Nicols is unsuitable. (*London B.Sc.*)

15. Give a description of the Laurent polarimeter and explain the action of its various parts. Discuss the merits of this particular type of polarimeter. (*London B.Sc.*)

16. Give a description of some form of polarimeter suitable for measuring the rotatory power of a sugar solution and explain the theory of its action. What do you consider to be the cause of the rotation produced by such a solution ? By what other means is it possible to rotate the plane of polarisation of light ? (*Tripos, Part I.*)

17. Give a short account of the contributions to the study of optics of the following : Snell, Römer, Huygens, Young, Fresnel. (*Camb. Schol.*)

18. Write a short essay on the polarisation of light, describing briefly methods of producing and detecting (a) plane, and (b) circularly, polarised light. Discuss the interference of light polarised in different planes. (*Camb. Schol.*)

19. Explain how plane polarised light may be converted into circularly polarised light, and how circularly polarised light may be distinguished from unpolarised light. Give the theory of any instrument used. (*London B.Sc.*)

20. Explain the action of (a) Nicol's prism, (b) a pile of plates, (c) a tourmaline in producing polarised light. How can circularly polarised light be produced ? (*London B.Sc.*)

21. What is meant by (a) plane polarised light, (b) circularly polarised light, (c) elliptically polarised light ? How may each of these be produced ?

How would you distinguish between

(a) unpolarised light and circularly polarised light,

(b) elliptically polarised light on the one hand and a mixture of plane polarised light and unpolarised light on the other ? (*Tripos, Part I.*)

22. Describe how polarised light may be produced and detected. Explain the colours seen when a parallel beam of light is passed through a Nicol prism, a piece of strained glass, and a second Nicol prism in series. (*Tripos, Part I.*)

23. Describe a method of measuring the velocity of light which can be set up over the short distance available in a laboratory.

Chapter XVI

THE EXTENSION OF THE SPECTRUM

161. THE INFRA-RED

We concluded Chapter V by stating two problems. We had replaced the physiological sensation of colour as a means of specifying different kinds of light by the physical quantity refractive index, but the refractive index of a given kind of light was different in different media. We required some kind of absolute refractive index, like an absolute temperature independent of the properties of a particular substance. We have already obtained the answer to this problem, for we have substituted for refractive index as specifying a given kind of light its wave-length, which is a number characteristic of the light and nothing else, if we use the wave-length in a vacuum. So the wave-length of light in a vacuum is our absolute refractive index ; it is the way of specifying the kind of light with which we are dealing. The second problem was this : if refractive index or wave-length, rather than physiological sensation, is the proper way of specifying the kind of light with which we are dealing, why should light stop at wave-lengths of 4×10^{-5} and $7{\cdot}5 \times 10^{-5}$ cm. ? It is obvious that any waves whose wave-lengths lie outside these limits will not affect the eye, as we know by experience, but that is no reason why such waves should not exist. We cannot call them light waves, if they do exist, but we shall call any waves in the ether **radiation**, using the term visible radiation to denote light.

To see if such waves do exist outside the limits of the visible spectrum, we must produce a spectrum and search for them beyond the red and violet respectively. But what detector shall we use when the eye fails ? Light produces two other effects besides the sensation of sight ; it produces heat, if it falls on a body which absorbs it, and it affects a photographic plate. In 1800 Herschel detected the presence of waves beyond the red end of the spectrum by their heating effect and they are called **infra-red radiation** from their position relative to the visible spectrum. He produced a spectrum in the usual way and placed the bulb of a thermometer, which had been blackened to make it absorb all the radiation falling on it, in the violet end of the spectrum. The rise in temperature was very small. As he moved it towards the red end, the rise in temperature increased and grew bigger when the thermometer moved beyond the red end of the spectrum. The maximum temperature

recorded by the thermometer occurred at a point some distance beyond the red end of the spectrum.

The best methods of detecting and measuring infra-red radiation depend on allowing it to fall on an instrument, which absorbs it and converts it into heat, which causes a rise in temperature in some mechanism very sensitive to temperature changes. The two main types of instrument, which are fully described in the author's Text Book on Heat, are the **thermo-pile** and the **bolometer**. The thermo-pile is essentially a thermo-couple, one junction of which is exposed to the radiation, the other being kept at a constant temperature. The radiation heats up the one junction and causes a current to flow in the circuit, whose magnitude is proportional to the amount of radiation falling on the junction in unit time. The bolometer is a strip of blackened platinum foil forming one of the arms of a Wheatstone Bridge. The bridge is balanced and the strip is then exposed to the radiation, which raises its temperature, alters its resistance, and upsets the balance of the bridge. The galvanometer deflection is proportional to the rate at which radiation falls on the bolometer. Finally infra-red radiation can be detected nowadays by suitable photographic plates. The original plate was sensitive to the blue and violet only ; then the orthochromatic plate sensitive to the green was produced, followed by the panchromatic plate, which is sensitive to the whole visible spectrum. This was produced by dyeing the emulsion with eosin, or erythrosin, or ethyl red. Finally plates sensitive to the infra-red were produced by dyeing them with a group of substances known as the cyanines. We shall not describe how the above methods of detecting and measuring infra-red radiation can be used to find how the number of ergs of radiation emitted per unit area in unit time varies with the temperature and nature of the source, as this problem is treated in the author's Text Book on Heat. We shall confine ourselves here to the precise nature of infra-red radiation and its relation to visible radiation.

The fact that infra-red radiation is emitted by the same sources as light, is refracted by a prism in a similar way, and produces heat in the same way, suggests that it is a wave motion in the ether like light. We should expect its wave-length to be rather longer than that of red light, just as red light has a greater wave-length than that of violet light. These views are confirmed by the facts that infra-red radiation obeys the same laws of reflection as light, it can be polarised, and finally it can be diffracted by a concave reflection grating (Art. 163), from which its wave-length can be measured and bears out the above expectation. So the only physical difference between light and infra-red radiation is merely one of wave-length ; the difference which seems so important at first sight, that it does not produce the sensation of sight, may be outstanding for the human being, but it is of little scientific interest ; the scientist has long since learned to regard physiological sensation with suspicion !

Infra-red radiation has found a new use in recent years in long-distance

photography. It is known that light is scattered to one side by dust in the atmosphere and by the molecules of the air itself, and the amplitude of the scattered light is inversely proportional to the fourth power of the wave-length. So blue light is scattered sixteen times as well as red light, which accounts for the red colour of the sun at sunset, when the sun's rays have passed through such a long length of atmosphere, and also for the blue sky, which is seen entirely by scattered light. Since the wave-length of infra-red radiation is even longer than that of red light, it will be scattered even less than red light, so that it will penetrate the atmosphere well. It is possible to take quite clear photographs of mountain scenes fifty miles away, using a filter passing only infra-red radiation and a suitable photographic plate.

162. INFRA-RED SPECTRA

We have already seen (Art. 40) that elements such as sodium and nitrogen, when suitably excited, give line or band spectra in the region visible to the human eye, and it is interesting to see if the spectra of such elements extends into the infra-red also. That is, if sodium is excited to give out its characteristic yellow light, does it, at the same time, give out any infra-red radiation ? If a spectrum is formed, will there be lines in the infra-red as well as the pair of lines in the yellow ? To settle this question, we need some kind of spectrometer for producing spectra in the infra-red as well as the visible. Light of different wave-lengths can be separated either by a prism or by a diffraction grating ; the prism will give lines of the greater intensity, since most of the radiation falling on a grating goes straight through and contributes nothing to the spectrum, while all the radiation of a given wave-length incident on the prism will go to form the corresponding line in the spectrum. But it is not possible to determine the wave-length of the lines formed by a prism spectrometer directly from their deviations, whereas the diffraction grating forms a spectrum the wave-lengths of whose lines can be simply calculated from their angles of diffraction. Since the sources of infra-red radiation are never very intense, it is important to get as great a proportion as possible of the incident light into the spectrum, so it is usual to employ a prism spectrometer in the infra-red, but it must be calibrated by a grating spectrometer using a source intense enough to get lines which are accurately detectable. It is interesting to note how much more sensitive the eye is than a thermo-pile ; the eye can detect 10 quanta of yellow light per second, while the thermo-pile needs 10^{14} of the same quanta per second to affect it. The photographic plate is superior to both in the sense that it responds to the *total* quantity of radiation received during the exposure, while the eye or thermo-pile responds to the *rate* at which radiation falls on them.

Certain difficulties arise at once in designing an infra-red spectrometer. If we decide to employ the type used for examining spectra in the visible, both its prism and lenses are made of glass. If a source of infra-red

radiation, such as an electric heater arranged to be just not red-hot, is placed in front of a thermo-pile and various screens are placed between the source and detector, it can be shown that glass absorbs the infra-red strongly. So it will be useless to make the lenses and prism of the spectrometer of glass. A similar experiment shows that quartz transmits infra-red well, rock-salt still better, and sylvine, the crystalline form of potassium chloride, best of all. Assuming that we have a spectrometer to measure the wave-lengths of the infra-red, such as will be described below, it can be shown that glass is transparent up to 30,000 Å.U., quartz to 40,000 Å.U., fluorspar or calcium fluoride to 90,000 Å.U., rock-salt to 150,000 Å.U., and sylvine to 230,000 Å.U. The Angstrom Unit, denoted conventionally by Å.U., is 10^{-8} cm. The prism of the spectrometer is accordingly made of quartz, fluorite, or rock-salt; it is best to use quartz, if it is transparent enough, as it has a much greater dispersive power than rock-salt or sylvine. If either of these two have to be used, they must be protected from moisture, whose absorption would spoil their crystal form, by dipping them into a solution of collodion in ether. The ether evaporates and leaves a thin film of collodion, transparent to the infra-red, over the prism. Lenses can be made of the same materials, but it is better to dispense with them altogether and to collimate the beam incident on the prism and focus the emergent rays with concave mirrors. This overcomes any difficulties with regard to absorption, it eliminates the fact that the focal length of the lens is different for different wave-lengths, and it makes it possible to focus the spectrum with visible light, when it will be correctly adjusted for any infra-red radiation, whatever its wave-length. A common type of infra-red spectrometer is illustrated diagrammatically in Fig. 255, in which the source of infra-red radiation is placed behind a slit S_1, from which rays diverge to the stainless steel concave mirror M_1. As the slit is placed at the focus of the mirror, the rays are reflected from it as a parallel beam and fall on the prism P, being refracted on to the stainless steel mirror M_2, which brings the parallel beam corresponding to each wave-length to a focus at a different place in the focal plane of the mirror M_2. The slit S_2 and the thermo-pile T are moved along the focal plane of the mirror in a direction at right angles to its axis, and a line in the spectrum is detected by the fact that the deflection of the galvanometer connected to the thermo-pile rises to a

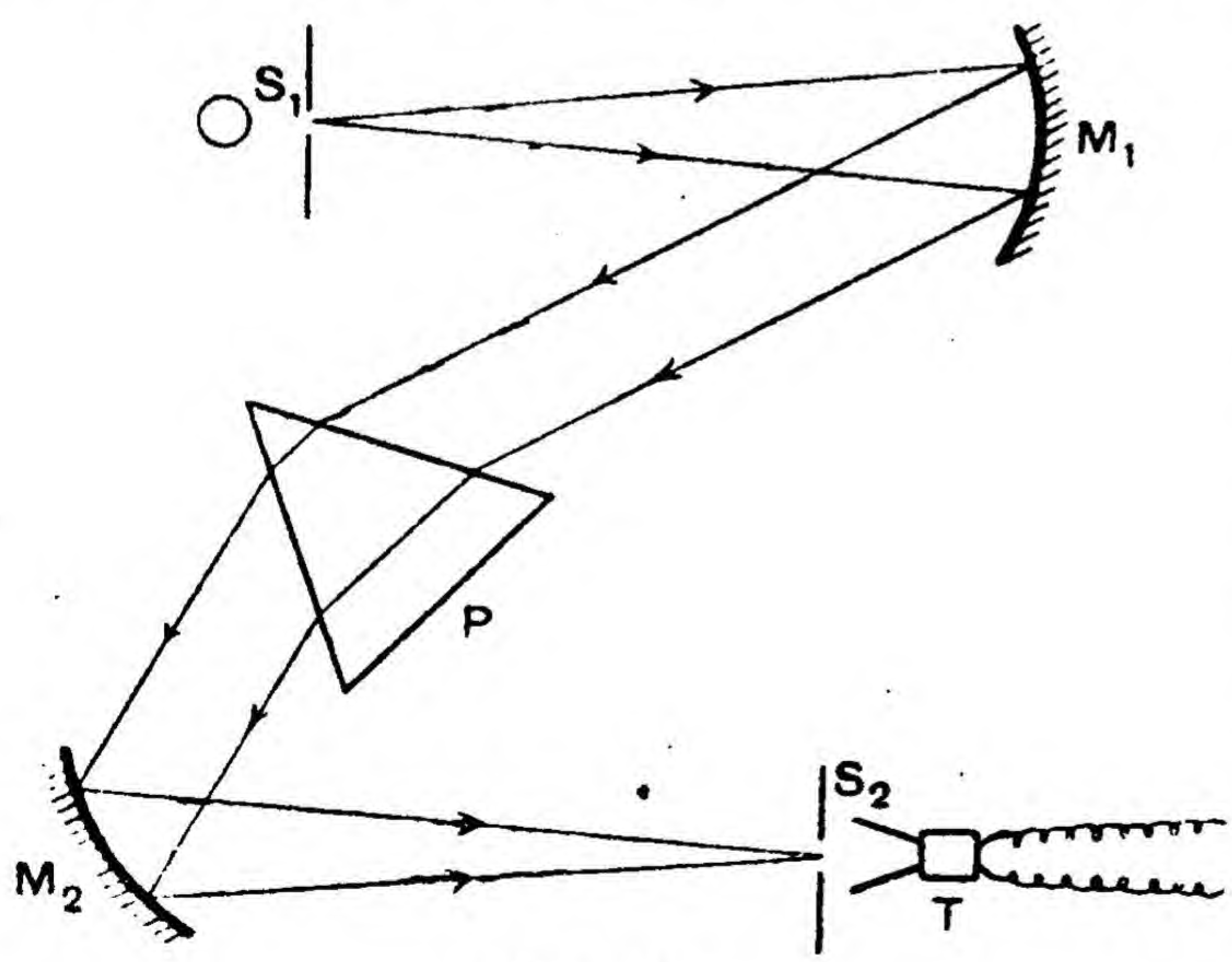


Fig. 255.

maximum when the line falls on the slit S_2. The spectrum of any source can be mapped out in this way and it may be repeated that the use of mirrors, which must be of stainless steel and not of glass, enables the spectrometer to be set up and adjusted with monochromatic visible light, which is very much simpler than doing it directly with infra-red radiation.

It is more convenient and more accurate to keep the thermo-pile or bolometer fixed in position and to bring the different lines in the infra-red spectrum on to the detector by rotating the prism. This is done in the **constant deviation spectrometer**, in which a special prism shown in Fig. 256 is used to deflect the selected line on to a fixed bolometer. The whole point of this spectrometer is that a plane mirror M is attached to the prism P as shown. If AB represents a ray of the incident radiation, BC is the ray of the particular wave-length which goes through the prism

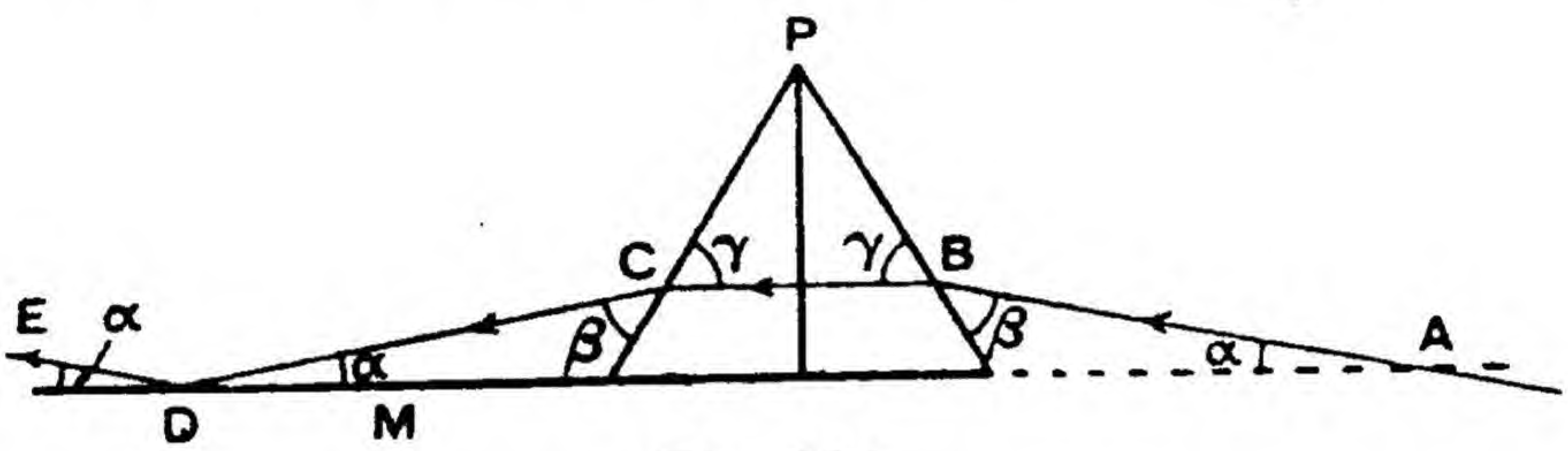


Fig. 256.

symmetrically and the emergent ray CD makes the same angle β with the prism as the incident ray. Therefore it follows from the simple geometry of the diagram that CD makes the same angle α with the mirror M as the incident ray AB makes with the base of the prism produced. Therefore the ray DE reflected off the mirror will be parallel to the original incident ray AB. This will be true whatever the angle α and hence this combination of prism and mirror ensures that the ray going through the prism at minimum deviation emerges from the mirror parallel to the direction of the incident ray; that is, the ray going through the prism at minimum deviation always comes from the mirror in the same direction. It is easy to arrange a concave mirror to focus this ray and all the others parallel to it on to a fixed bolometer. But, for a given angle of incidence of the ray incident on the first face of the prism, only one particular wave-length passes through at minimum deviation and falls on the bolometer. The different wave-lengths are brought on to the bolometer in turn by altering this angle of incidence by turning the prism about its refracting edge as axis. If a source of infra-red radiation is used, whose wave-lengths have already been measured by a reflection grating, the reading of the vernier on the prism turntable can be taken for each known wave-length and a calibration curve of wave-length against the reading of the prism turntable can be drawn up. A very sensitive type of vernier must be used for this purpose, since the dispersion of a prism, which decreases as we go from violet to red and red to infra-red, is so small in the infra-red that the whole infra-red is included in 10° with a typical

prism. Once the above calibration has been done, the spectrometer can be used to determine the wave-lengths of the infra-red lines of any source. If the radiation from a sodium lamp is examined in such a spectrometer its spectrum will be found to contain not only the doublet in the yellow, but also a large number of doublets in the infra-red.

163. THE CONCAVE REFLECTION GRATING

We shall now describe how the wave-length of infra-red lines can be measured by the diffraction grating. The usual transmission grating of glass is useless since glass absorbs the infra-red, but it is possible to make diffraction gratings, in which the reflected light is diffracted on account of the very small width of the reflecting surface. They can be made by ruling a set of parallel equidistant lines on a plain piece of speculum metal with a diamond ruling point, the narrow lines of untouched metal between the rulings being the parts of the metal from which the light is reflected, most of the light taking the direction obeying the law of reflection, but some being diffracted in directions on either side of this. Let A, B, C, D, . . . (Fig. 257) represent the smooth parts of the metal mirror from which reflection takes place, the spaces in between representing the rulings produced by the diamond point, from which no regular reflection of any kind can occur. If a parallel beam is incident at an angle i, consider the radiation diffracted at an angle θ with the normal to the grating. The two rays striking the lines A and B are in phase at A and F,

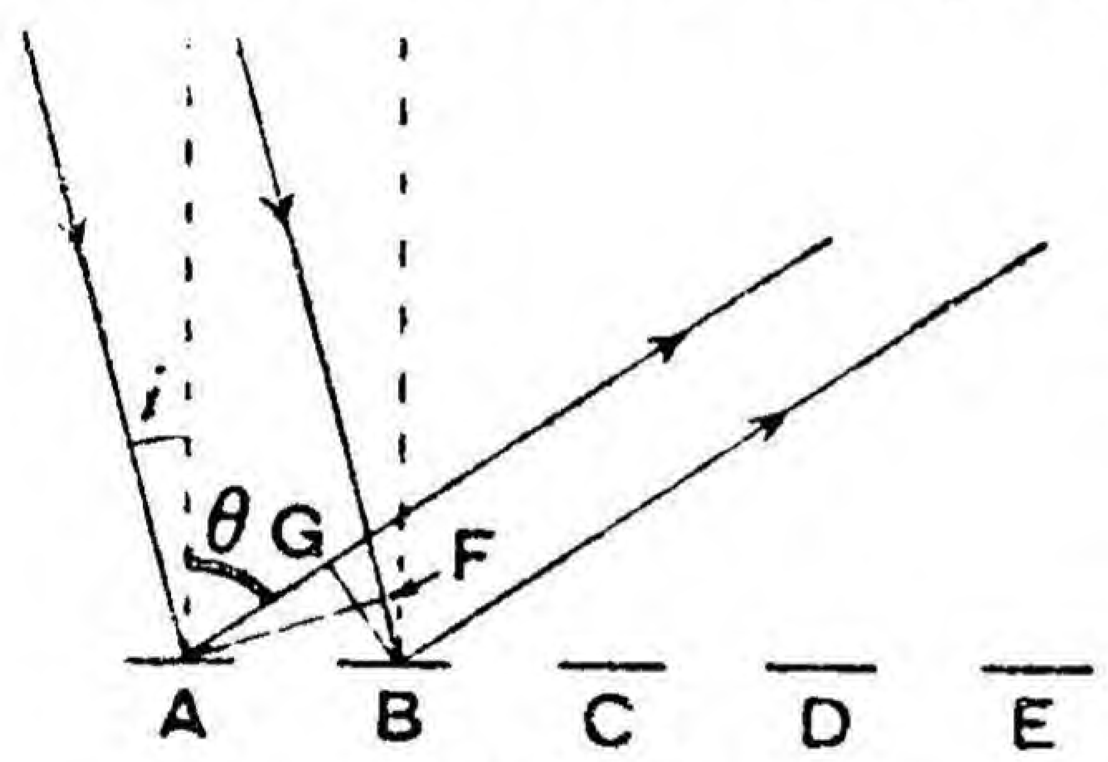


Fig. 257.

while the rays leaving those lines at an angle θ to the normal must be in phase at B and G, if there is to be a maximum in this direction. For, if two crests start from B and G simultaneously, they will arrive at the same time at the focus of a lens, whose axis makes an angle θ with the normal to the grating, and will reinforce each other. Hence the path difference between the rays falling on successive lines is AG—BF or $e(\sin\theta-\sin i)$. So we get maxima in those directions given by the equation

$$e(\sin\theta-\sin i)=n\lambda$$

where n is an integer.

This grating could be used for the infra-red, but it would still need quartz or rock-salt lenses both to collimate the incident light and to focus the light diffracted from the grating. But Rowland saw how to avoid the use of lenses by ruling the lines on the surface of a concave mirror, the lines being such that their projections on a flat surface are parallel and equidistant. The ruled mirror both diffracts the radiation and brings it to a focus. The grating is about 6 in. wide and its radius of curvature is some 10 ft. If the source of radiation A (Fig. 258) is placed on a circle

whose diameter is equal to the radius of curvature of the grating, the spectra of the various orders will be in focus along an arc of the same circle. Let PQ be the grating and C be its centre of curvature; the width of the grating is so small compared with its radius of curvature that the surface of the grating coincides with the circle on which the source and the spectra lie. Consider the ray AP incident at an angle i, which gives rise to a ray PB diffracted at an angle θ. Draw the lines AQ and QB. Then $\angle AQC = \angle APC$, since they stand on the same arc, AC and $\angle CQB = \angle CPB$ for the same reason. Hence the rays AQ and QB have the same angles of incidence and diffraction respectively as AP and PB. Therefore, if the angle of diffraction θ is a maximum for a wave-length λ incident at an angle i, all the rays starting from A and diffracted at that angle at any point of the grating will reinforce one another at B; in fact, a line of wave-length λ in the spectrum is formed at B. Similarly lines corresponding to other wave-lengths are formed on the arc of the circle in the neighbourhood of B. So, if A is a source of radiation, the spectra of different orders will be in focus on the circle through A, B, and C whose diameter is equal to the radius of the grating.

Finally Rowland designed an ingenious mounting for the grating, source, and photographic plate or detector illustrated in Fig. 259. The

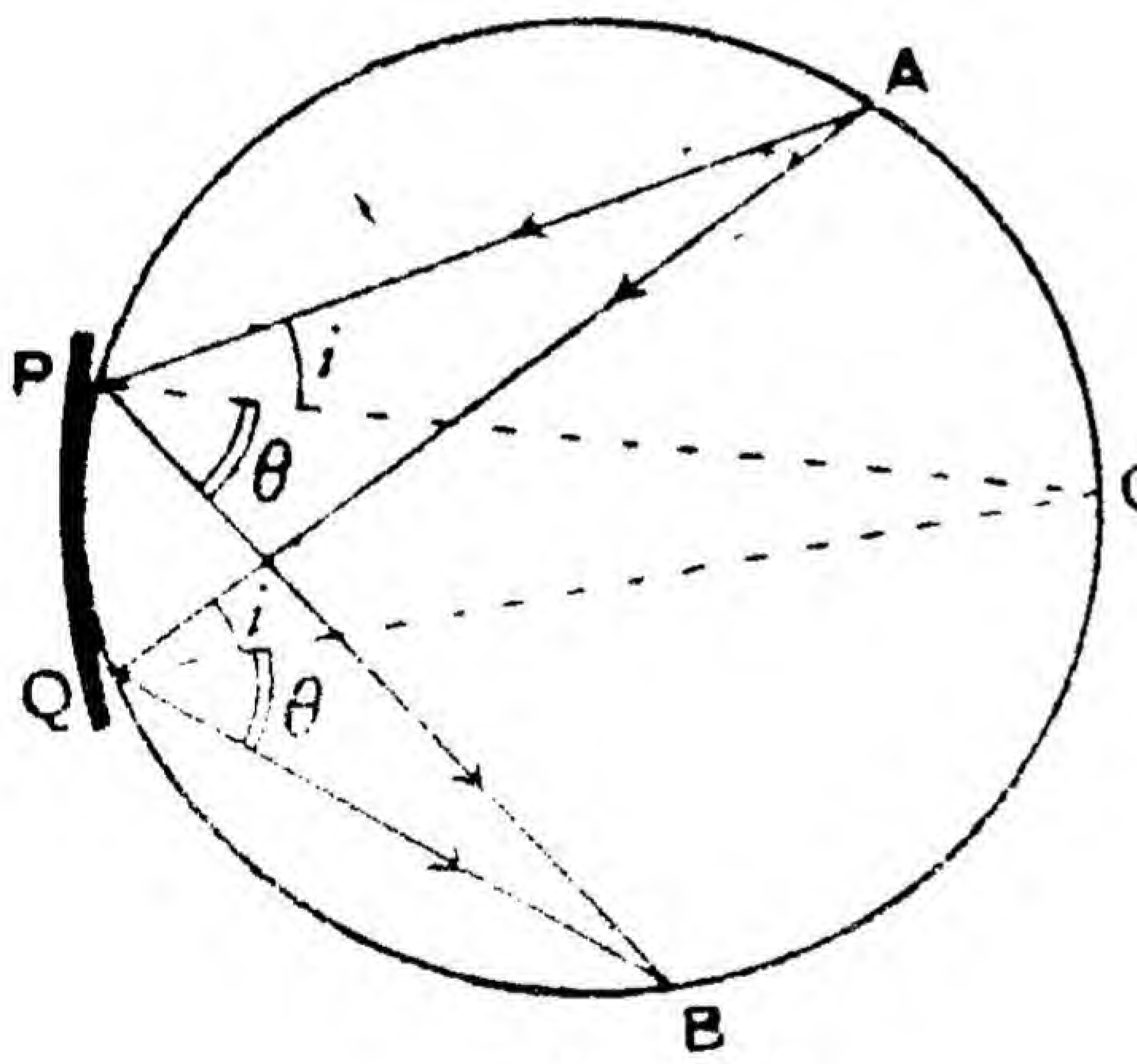


Fig. 258.

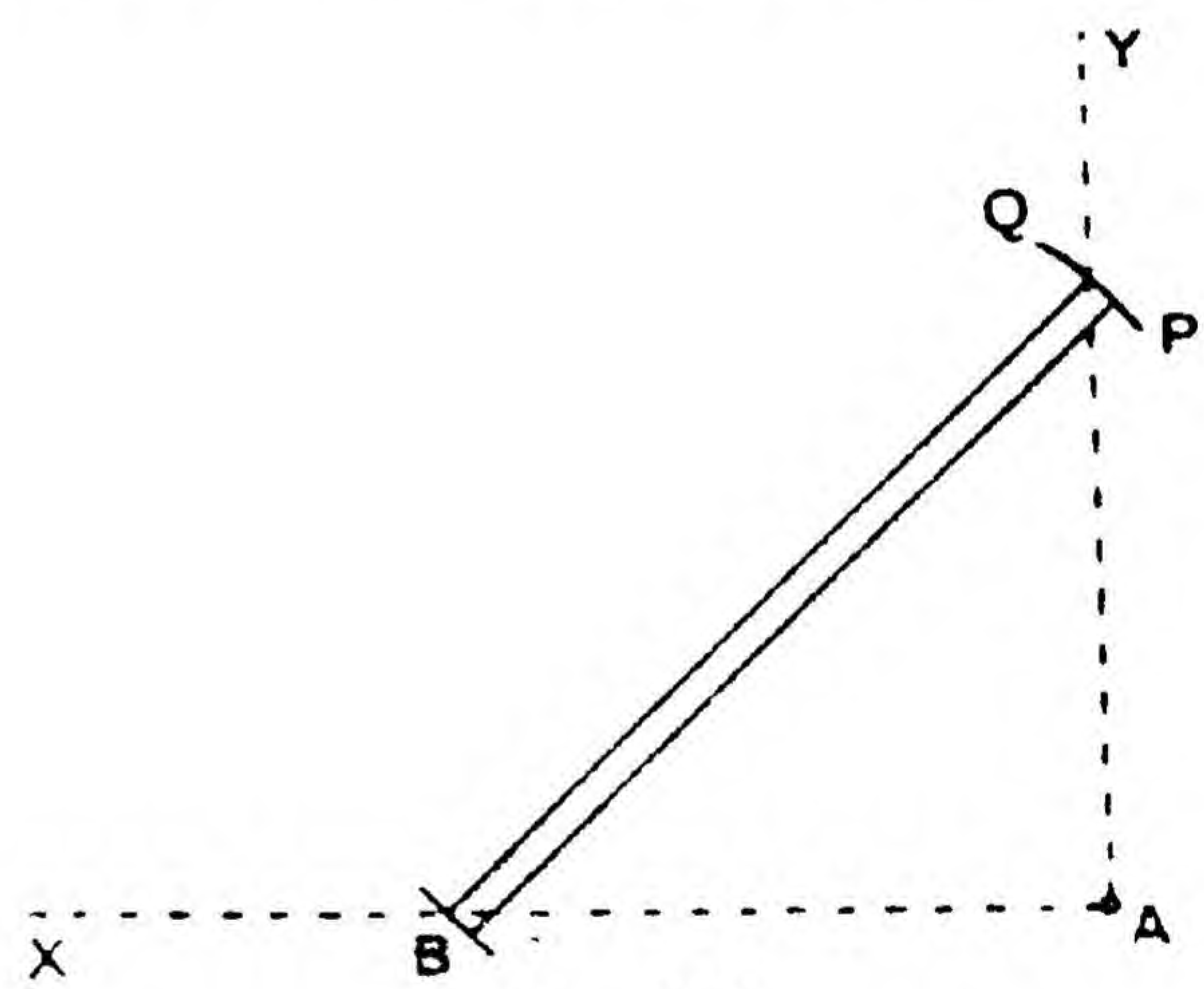


Fig. 259.

grating PQ and the plate B are mounted on the ends of a steel arm, whose length is equal to the radius of curvature of the grating. This arm slides up and down the rails AX and AY perpendicular to one another, the source of radiation being situated at A. Consequently the source, grating, and receiver all lie on the circle whose diameter is equal to the radius of curvature of the grating, so that the rays of a given wave-length from A will come to a focus at B, if the angle of incidence is suitably adjusted. This is done by moving the wooden arm into the correct position, when the different wave-lengths come to a focus at B in turn. The Rowland concave grating is the standard instrument for measuring wave-lengths both in the infra-red and ultra-violet, since it eliminates the use of lenses. It has the serious weakness that only a small proportion

of the incident radiation goes into the diffracted spectra, which are therefore weak. Consequently long exposures are necessary to get the spectral lines bright enough to be measured accurately and the grating must be used in a basement at constant temperature, since change in temperature shifts the lines and blurrs them. The Rowland grating is used largely for measuring standard wave-lengths, and these are used for calibrating the prism spectroscopes, which give much stronger spectra and can be used to investigate sources of small intensity.

164. INFRA-RED RADIATION AND WIRELESS WAVES

Infra-red radiation of wave-length up to 230,000 Å.U. or 0·0023 cm. can be produced and detected by the use of sylvine prisms. In 1888, however, another kind of radiation of still longer wave-length was found by Hertz, when he discovered the electric waves, which are now used to transmit speech and music in broadcasting. These electric waves, or wireless waves as they are often called, obey the same laws of reflection and refraction as light; like light, they travel through empty space with a velocity of 3×10^{10} cm. per sec. They also show interference and diffraction and they can be polarised. There can be no doubt that they are waves in the ether, differing only from light in wave-length. They are produced by setting electricity in oscillation by discharging a condenser through an inductance, the oscillations of the electricity setting up waves in free space. After twenty years it was possible to produce wireless waves with wave-lengths as big as 15,000 metres and as small as 100 metres. The question arises: is there a gap in the spectrum between the infra-red radiation and the wireless waves or is it continuous?

Attempts were made to close the gap from both ends by seeking longer infra-red radiation and shorter wireless waves. Longer infra-red radiation was found in a curious way. The reflecting power of quartz for different wave-lengths was being investigated, when it was found that quartz was a perfect reflector for 85,000 Å.U., 96,000 Å.U., and 207,500 Å.U. It was natural to extend this investigation to other crystals such as rock-salt and sylvine, when it was found that the former was a perfect reflector for 522,000 Å.U. and sylvine for 614,000 Å.U.. Quartz was then found to be transparent to these long infra-red waves with an abnormally high refractive index, 2·2, and this led to even longer waves being isolated. A Welsbach incandescent mantle W (Fig. 260) was placed behind a hole H_1 in a screen to act as the source of radiation and a quartz lens L was placed at a distance $2f$ from the illuminated hole, where f is the focal length of the lens for the long infra-red waves. Another screen with a hole H_2 is placed at a distance $2f$ on the other side of the lens, so that long infra-red waves emerging from the first hole will pass through the second one. But shorter waves and visible light, whose refractive

index is *less* than that of the long waves, will not converge to the same extent and may even diverge, and so they cover a wider area of the second screen. Consequently only the long infra-red waves will emerge from the second hole, provided that no shorter waves are allowed to emerge from it by stopping all the central rays with the stop P placed behind the lens. In this way infra-red radiation of wave-length 1,070,000 Å.U., or

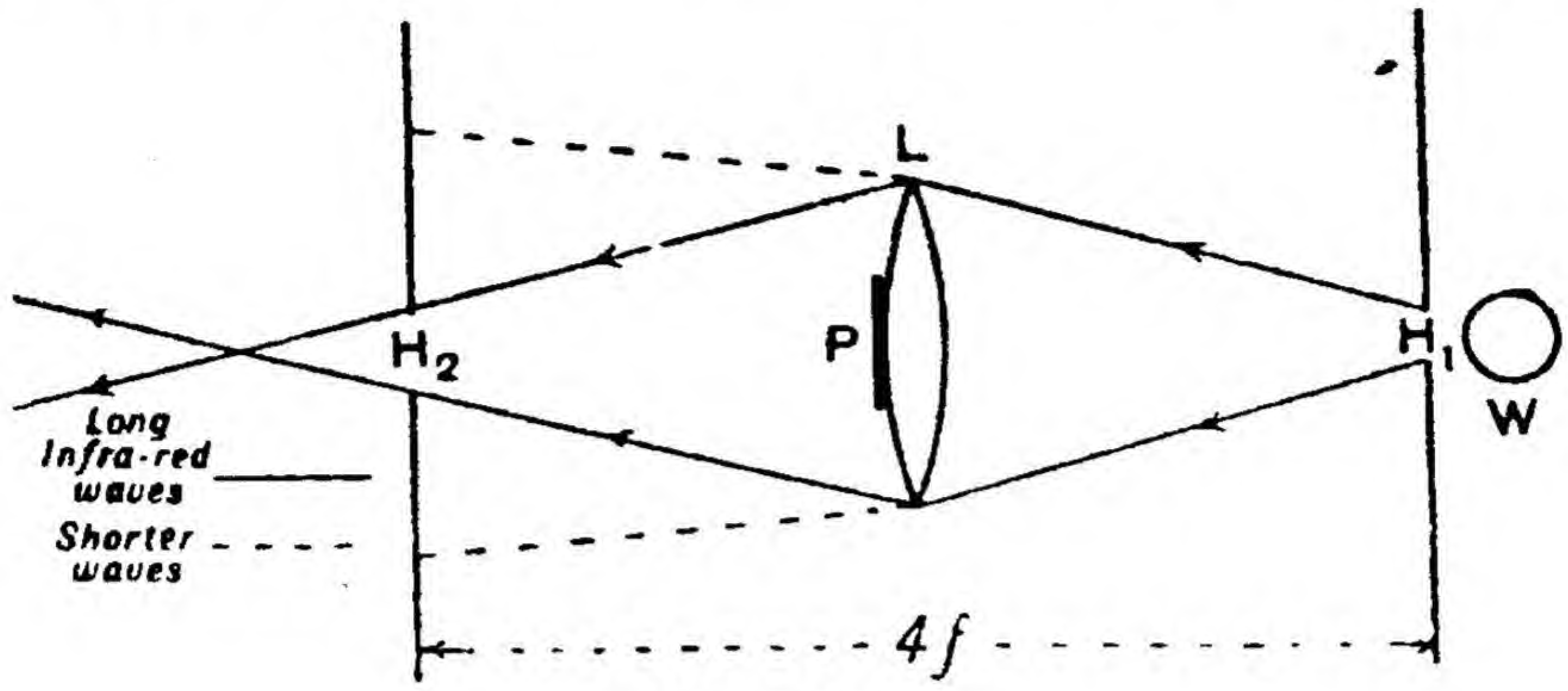


Fig. 260.

just over 0·1 mm., has been produced. It is detected by its heating effect and its wave-length is measured by a diffraction grating consisting of wires about 0·1 mm. apart wound on a frame. The production of very short wireless waves reduces itself to obtaining an oscillating circuit in which both the capacity and inductance are very small, and short waves of 0·1 mm. have now been produced, which are detected by their electrical effects and whose wave-lengths are measured by electrical means. These wireless waves can also be detected by their heating effect and so there is no doubt as to the continuity of this region of the spectrum.

165. THE ULTRA-VIOLET

Having extended the spectrum from the red end of wave-length 7500 Å.U. to a wireless wave of wave-length up to 30,000 metres, it is now time to turn our attention to the other end of the spectrum. We must retrace our steps and return to the year 1801, when we find Ritter observing that, if a photograph of a spectrum is taken, the plate is affected beyond the violet end of the visible spectrum. This shows that the source of radiation contains rays which are refracted like light, but whose refractive index is rather greater than that of violet light. This is called ultra-violet light, or better, ultra-violet radiation. It is reflected in the same way as light, but it is interesting that silver reflects only 4 per cent. of the ultra-violet falling on it ; indeed a thin film of silver forms a very good filter for transmitting the ultra-violet while stopping the visible ! The best reflecting surface for ultra-violet radiation is a surface of aluminium on glass or quartz. In 1811 Young photographed Newton's Rings in ultra-violet radiation, thus confirming their identity with light. The radii of corresponding rings was less with ultra-violet radiation than with violet light, showing that the wave-length of ultra-violet radiation is

less than that of light, which is to be expected from its greater deviation by a prism. The fact that ultra-violet radiation is waves in the ether differing from light only by its smaller wave-length was finally confirmed by measuring its wave-length with a reflection grating.

Ultra-violet radiation can be detected in four ways; photographic effect, fluorescence, phosphorescence, and photo-electric effect. The first method has already been referred to. The second consists in the fact that, if certain substances are placed in a beam of ultra-violet radiation, they emit visible light. For example, a solution of quinine sulphate in dilute sulphuric acid gives blue light, a crystal of fluorspar (calcium fluoride) gives out violet light, uranium oxide gives out green light as does barium platino-cyanide, and so on. **The wave-length of the light emitted by a fluorescent substance is usually greater than that of the exciting radiation**, which is known as **Stokes' law**, but there are exceptions to it. The third way of detecting ultra-violet radiation is to place some calcium sulphide in a beam of the radiation, when it will continue to glow with a bright light or phosphoresce for as long as an hour after the radiation is switched off. This is the way in which the luminous hands and figures of a watch work. Phosphorescence can also be obtained with barium and strontium sulphides. Finally, ultra-violet light can be detected by the fact that it causes an emission of electrons when it falls on a metal plate, the photo-electric effect.

The first two of these four methods are the commonest ones for the detection of the ultra-violet and it is easy to show that glass absorbs the ultra-violet strongly, being opaque to wave-lengths below 3000 Å.U. Quartz is transparent down to 1800 Å.U. and fluorspar to 1200 Å.U. Prism spectrometers must use lenses and a prism of quartz or fluorspar and the concave grating is the best instrument for the direct measurement of wave-lengths in the ultra-violet.

Ultra-violet radiation has been found to be a necessary element in the production of vitamin D, which is essential to healthy life and it is being used in medical treatment nowadays. It is common to give a dose of ultra-violet radiation to miners, whose occupation below the ground deprives them of the normal amount of ultra-violet radiation, which comes from the sun. It has been found useful in the treatment of rickets and tubercular abscesses and diseases of the bone. It also has important uses in microscopy. The reader will recall that what we demand of an optical instrument is resolving power (Art. 143) and he will remember that the resolving power of a telescope is $\frac{1 \cdot 22\lambda}{D}$, where λ is the wave-length of the radiation and D is the diameter of the objective of the telescope. It follows that the resolving power of a telescope would be greater if ultra-violet radiation were substituted for visible radiation. The same thing is true of a microscope. It is easy to see the physical principle lying behind this result. An object is seen or rendered visible by the effect it produces on the

radiation which falls on it. Detail is only shown up in an object, when the size of the detail is of the order of the wave-length of the radiation falling on it. If the size of the grain is less than the wave-length of the radiation, it cannot produce any effect on the radiation, just as the long waves of the sea pass over a small projecting rock and a few yards later the two halves of the wave close up bearing no trace of the presence of the rock. Consequently, if a smaller wave is used, grains of smaller size can be distinguished in the object, that is, more detail can be seen. This is precisely what the use of ultra-violet radiation in microscopy has done, the eye being replaced by a photographic plate. Photographs of bacteria showing greater detail than has been obtained with the visible have been taken in ultra-violet radiation.

166. SOURCES OF RADIATION

Before describing how spectra in the ultra-violet are investigated, we shall give an account of the various sources of radiation, including the ways in which elements may be made to give out their characteristic spectra. The most copious source of ultra-violet radiation is the carbon arc, in which a current of several amperes is made to flow between two carbon rods by a P.D. of the order of 100 volts. The positive carbon becomes hollowed out and is the source of most of the radiation, while the negative carbon becomes pointed. The radiation is caused by the particles of carbon being raised to white heat due to the heat developed by the passage of the electric current. If the crater of the positive carbon is filled with an element or one of its compounds, the flame of the arc will emit the **arc spectrum** of the element. The arc spectrum of sodium and mercury can also be obtained from the sodium or mercury vapour lamps described in Art. 88. These lamps have the same electrical characteristics as the carbon arcs, in that the current and P.D. are of the same order and the mercury vapour lamp is particularly rich in ultra-violet radiation.

The **flame spectrum** of an element is excited by putting the element or one of its compounds in the hot part of a bunsen flame, where it is vapourised. The flame spectrum of sodium, for example, is produced by soaking a piece of asbestos in a solution of common salt and putting the asbestos into a bunsen flame, when the flame gives out the characteristic yellow light, which is analysed into two lines close together in the yellow in a spectrometer. The **spark spectrum** of a metal can be produced by connecting two rods MM of the metal (Fig. 261) to the secondary of an induction coil and putting a condenser S in parallel with the spark gap. When the induction coil is started up, brilliant sparks pass between the metal rods and the light emitted is characteristic of the metal

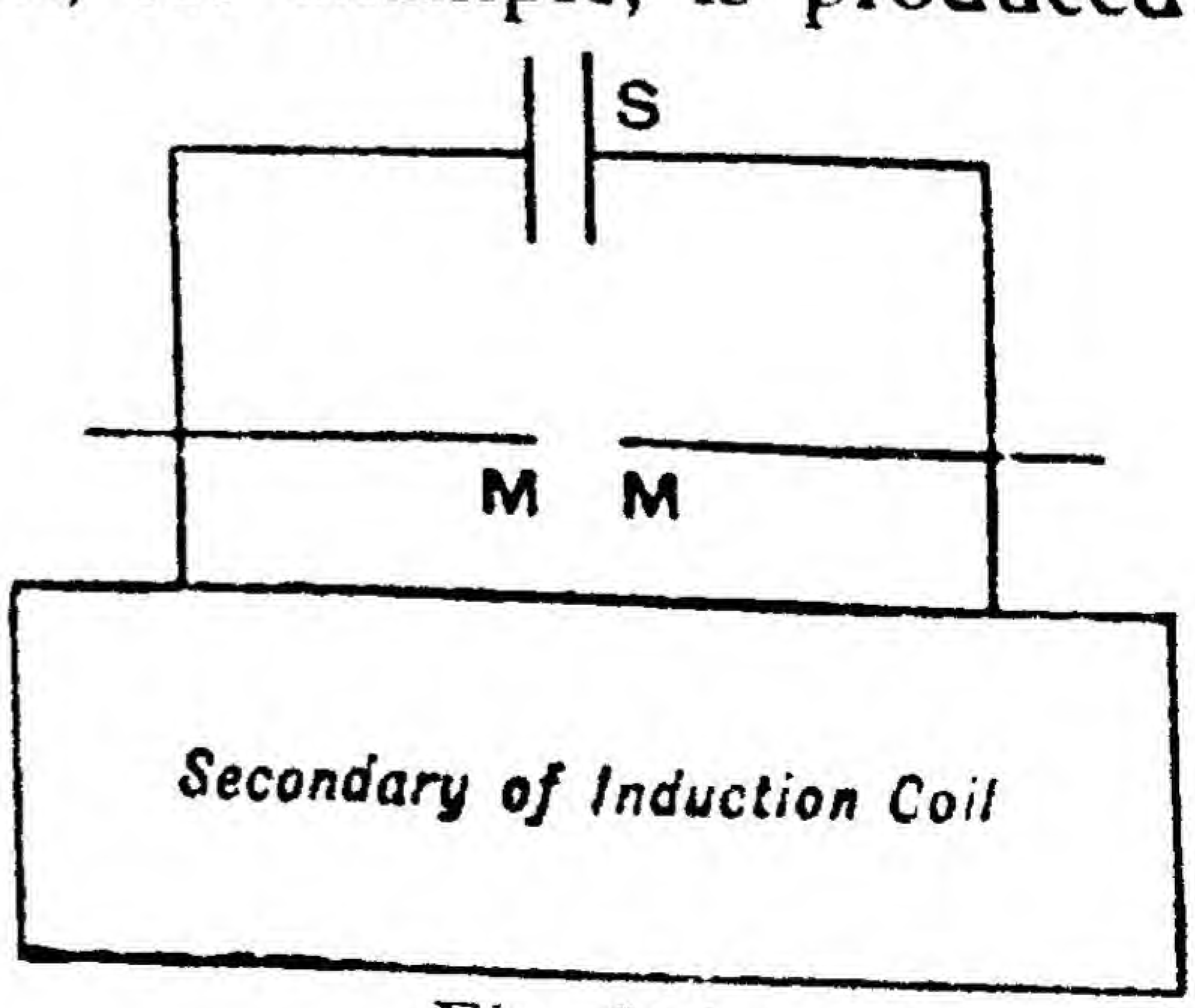


Fig. 261.

and the method of excitation. The resulting spectrum is called the spark spectrum of the metal. A fourth way of exciting an element to emit light is to exhaust a tube and fill it with the gas under test, such as hydrogen, to a pressure of 2 or 3 millimetres of mercury. If two electrodes are sealed into the tube and are connected to the secondary of an induction coil giving about 10,000 volts, an electric current will pass through the gas, which emits a characteristic light. This is a discharge tube (Fig. 262)

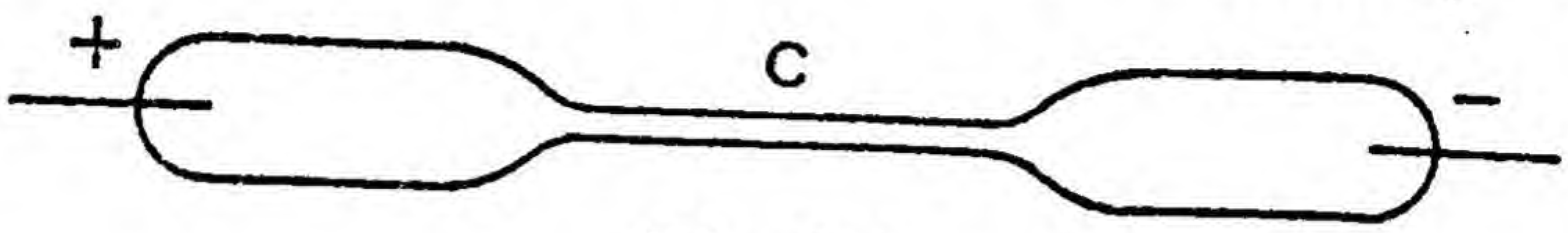


Fig. 262.

and is used in night signs for advertising purposes and for outlining buildings at night. The contrast between the conditions in this tube and the carbon arc or sodium lamp should be noticed; here the temperature is comparatively low, the P.D. is of the order of 10,000 volts, and the current is of the order of a few milliamps. The light is brightest in the capillary tube C joining the two electrodes and this tube forms a convenient line source of radiation, which can be focussed on the slit of a spectrometer.

167. ULTRA-VIOLET SPECTROMETERS

Wave-lengths of lines in the ultra-violet must be measured in the first place by a concave reflection grating, which is used in just the same way as in the infra-red. But the reflection grating has the disadvantage that most of the radiation is reflected according to the usual law and little of it goes into the diffracted spectra, which are therefore weak. Another factor contributing to this weakness is the fact that metals reflect ultra-violet radiation badly. The prism spectrometer is therefore to be preferred, because it gives stronger spectra. It must be calibrated in wave-lengths with a source of ultra-violet radiation, the wave-length of whose lines have been measured by a reflection grating. For this purpose the iron arc, a source rich in lines in the ultra-violet, is used, and the diffraction spectra are made sufficiently intense to be accurately measurable by a long exposure of the photographic plate to the radiation. When the prism spectrometer has been calibrated, it can then be used to measure the wave-lengths of the lines of less intense sources.

There are certain physical principles of interest in the design of an ultra-violet spectrograph, a conventional type being illustrated diagrammatically in Fig. 263. The prism must be of quartz, or fluorspar for the extreme ultra-violet; mirrors are not used for focussing, as silver reflects ultra-violet radiation so feebly. The focussing must be done by quartz or fluorspar lenses. If quartz is used certain difficulties arise. Quartz is doubly refracting, and to eliminate this the prism is cut with the optic axis parallel to the base of the prism, so that the ray traversing the prism

at minimum deviation shall suffer no double refraction. But quartz rotates the vibration plane of plane polarised light travelling along the optic axis and, if ordinary light travels through quartz in this direction, it is split up into two circularly polarised components vibrating in opposite senses and travelling with different velocities. Therefore the two components have slightly different refractive indices and two spectra nearly coinciding will be produced, which will decrease the sharpness of the spectral lines and the resolving power of the spectrometer. This is eliminated by the **Cornu prism**, which consists of two 30° prisms, one of left-handed and the other of right-handed quartz, while each lens is cut with its

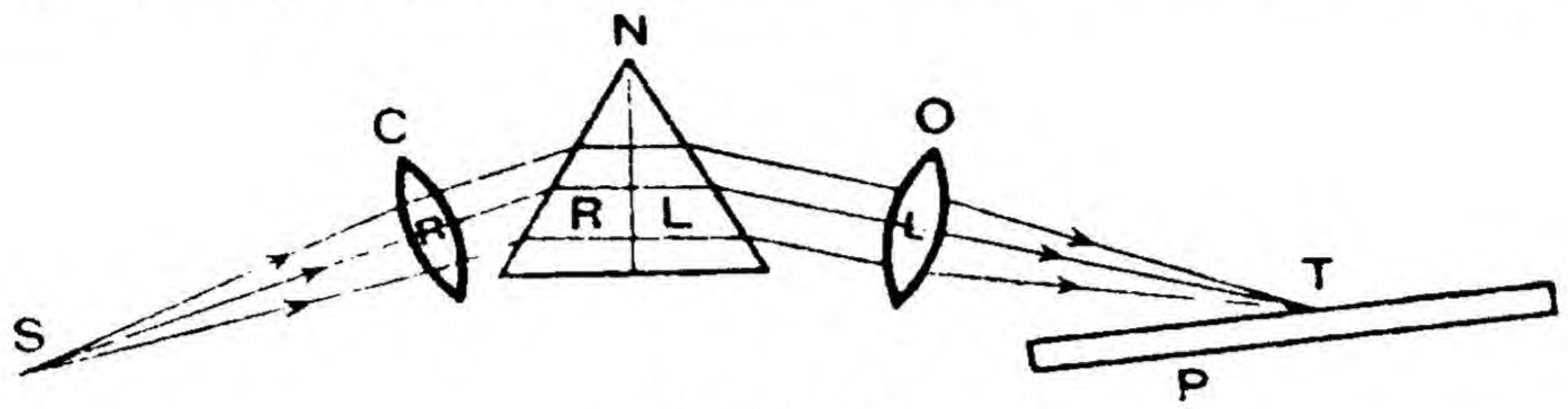


Fig. 263.

axis parallel to the optic axis of the quartz, one being left-handed and the other right-handed. The slight separation of the two beams produced by the first section of the prism is annulled by the second one and the same thing is true of the two lenses. Therefore this arrangement of lenses and prisms ensures that the ray going through the prism at minimum deviation shall not suffer any splitting whatsoever and therefore the corresponding spectral line has the maximum sharpness. One other adjustment must be made to get the spectrum in focus all over the photographic plate P. The dispersion of glass is greater in the violet than in the red, and the same thing is true of quartz ; also the dispersion in the ultra-violet is even greater than in the violet. It is so big, in fact, that the focal lengths of the lenses for the different wave-lengths vary so much that it is necessary to tilt the photographic plate to be at about 20° to the axis of the lens O, in order that a wide range of wave-lengths may come to a focus on the plate. But the lines will be sharpest in the middle, since all rays except the one going through at minimum deviation suffer a small amount of double refraction. The spectrum is focussed by replacing the plate by a fluorescent screen and setting this in such a position that the line going through at minimum deviation is sharply focussed and then the plate is put in this position and is ready for use. If the spectrometer has already been calibrated, when the spectra have been photographed the wave-lengths of the various lines can be read off from their positions on the plate.

An ingenious way of ensuring that the elimination of any doubling of the beam is as perfect as possible is achieved in the Littrow mounting illustrated in Fig. 264, in which only the central ray of the beam passing through the prism at minimum deviation is shown. The rays from the source S are reflected through a right angle by a plane aluminised

mirror and are made parallel by the lens L, after which they fall on the quartz prism N, whose back surface is coated with cadmium amalgam to make it reflecting. The wave-length which strikes the back surface

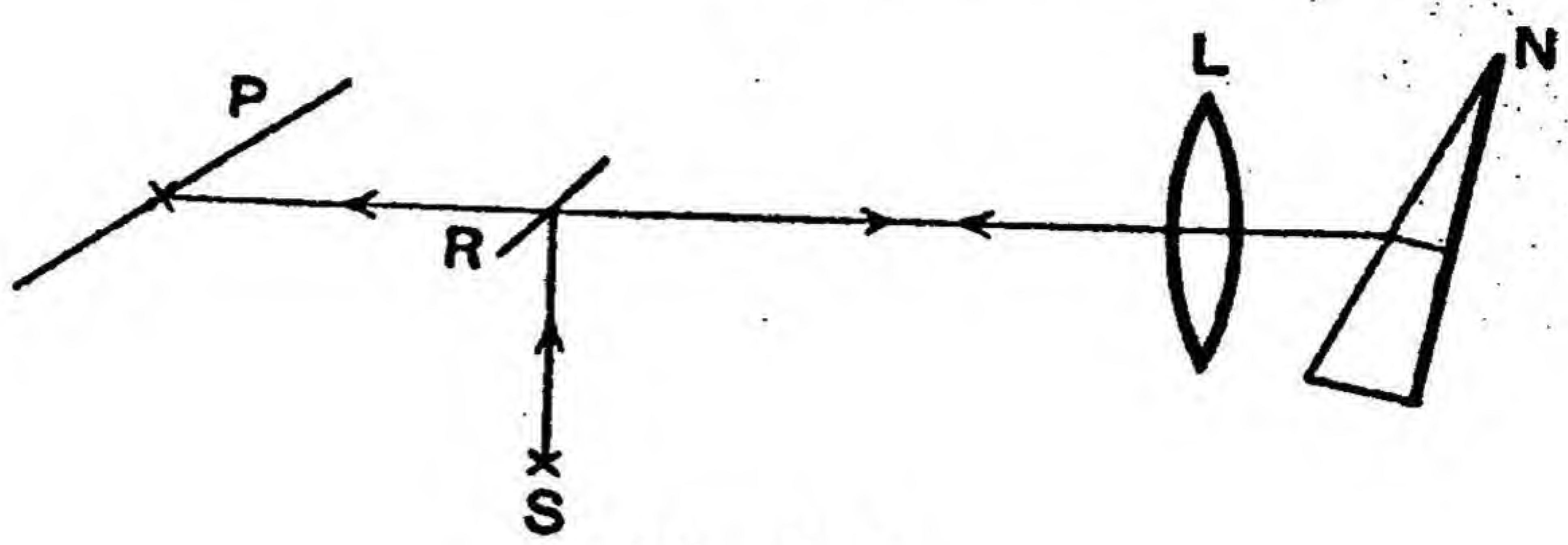


Fig. 264.

normally is reflected back along its path and, in addition, it goes through the prism at minimum deviation. Since the optic axis of the quartz is normal to the back face of the prism this ray suffers no double refraction, and any splitting due to the resolution of the light into two oppositely circularly polarised components is annulled when the ray retraces its path on the way out of the prism again. *The compensation is perfect in this case*, which is never quite true of the Cornu prism, as it is impossible to get two prisms of identical dimensions, one made of left-handed and the other of right-handed quartz. The beam then passes through the lens L again and is brought to a focus at a point on the plate P, which is the same distance away from the mirror R as the source S. The return ray is shown coinciding with the incident ray in the diagram, but, in practice, the mirror R is placed a little above the axis of the lens L, so that the return beam is thrown below the mirror and is not reflected back to C but goes straight on to the plate. The collimation is also perfect for all wave-lengths, since the same lens is used in both the collimator and the telescope. The Littrow mounting has the additional advantage that it is rather more compact than the more usual type of spectrometer.

The above spectrometers enable spectra down to 1200 Å.U. to be investigated ; Schumann and Lyman have got down to even shorter wave-lengths by using plates with a very thin layer of gelatine, so that the radiation was not absorbed before it reached the silver bromide, and also by exhausting the whole spectrometer so as to eliminate the absorption of the radiation by the air.

168. X-RAYS

We have seen that the spectrum has been detected on the short wave-length side of the visible region down as far as 1000 Å.U. and we must now pick up another clue from quite a different range of phenomena, which will lead to a still greater extension of the spectrum. A large number of experiments were being done on the passage of electricity through gases at low pressure towards the end of the nineteenth century and such work was one of the great centres of interest in physics, much as the disintegra-

tion of the atom is one of the centres of interest to-day. Cathode rays, light blue streamers emitted from the cathode of the discharge tube, had been discovered, and their nature was the subject of lively controversy. The English school of physicists held them to be negatively charged electricity, while the German school contended that they were yet another kind of radiation or wave motion. Perhaps the most recent experiments on these rays (Art. 178) show that there is an element of truth in each of these points of view. The chief point of importance for us, however, is that a large number of physicists were working with discharge tubes, of which a typical pattern is shown in Fig. 265. The tube has two electrodes sealed into it, C being the cathode and A the anode. The pressure in the tube must be lowered to about $\frac{1}{50}$ mm. of mercury in order to produce the cathode rays and at this order of pressure there is little else to be seen but the light-blue colour due to their passage along the tube. In 1895 Röntgen made a remarkable discovery when working with one of these tubes. He was looking for invisible radiation outside the tube to see if the second view of the nature of cathode rays was correct, and he covered the tube in black paper to keep any visible light from escaping out of the tube. He happened to have a screen of fluorescent material lying on the table near to the tube, and he was surprised to notice that it glowed brightly when he worked the tube by applying a P.D. of some 10,000 volts from the secondary of an induction soil to the electrodes. He also noticed that a box of photographic plates wrapped in the black paper lying near the tube was fogged, and on deliberately repeating this experiment he found that they were fogged again. This fogging must have been due to the discharge tube, since some of the plates were removed before exposure to the tube and were found to be unaffected. Although Röntgen was looking for something quite different, he felt that this clue was important and he decided to follow it up. So far he had discovered that the discharge tube emitted something which could penetrate glass and black paper, affected a photographic plate, and produced fluorescence. He called this " thing " X-rays. It was soon found that they were absorbed strongly by lead and that they cast sharp shadows, showing that they travelled in straight lines. It was then a simple matter to place a lead screen with four holes in it below the discharge tube and to place a photographic plate on the side of the screen remote from the discharge tube. On working the tube and developing the plate, there were four spots on it where the beam of X-rays had come through the holes. If the plate were replaced in position and the lines joining the spots on the plate to their respective holes were produced

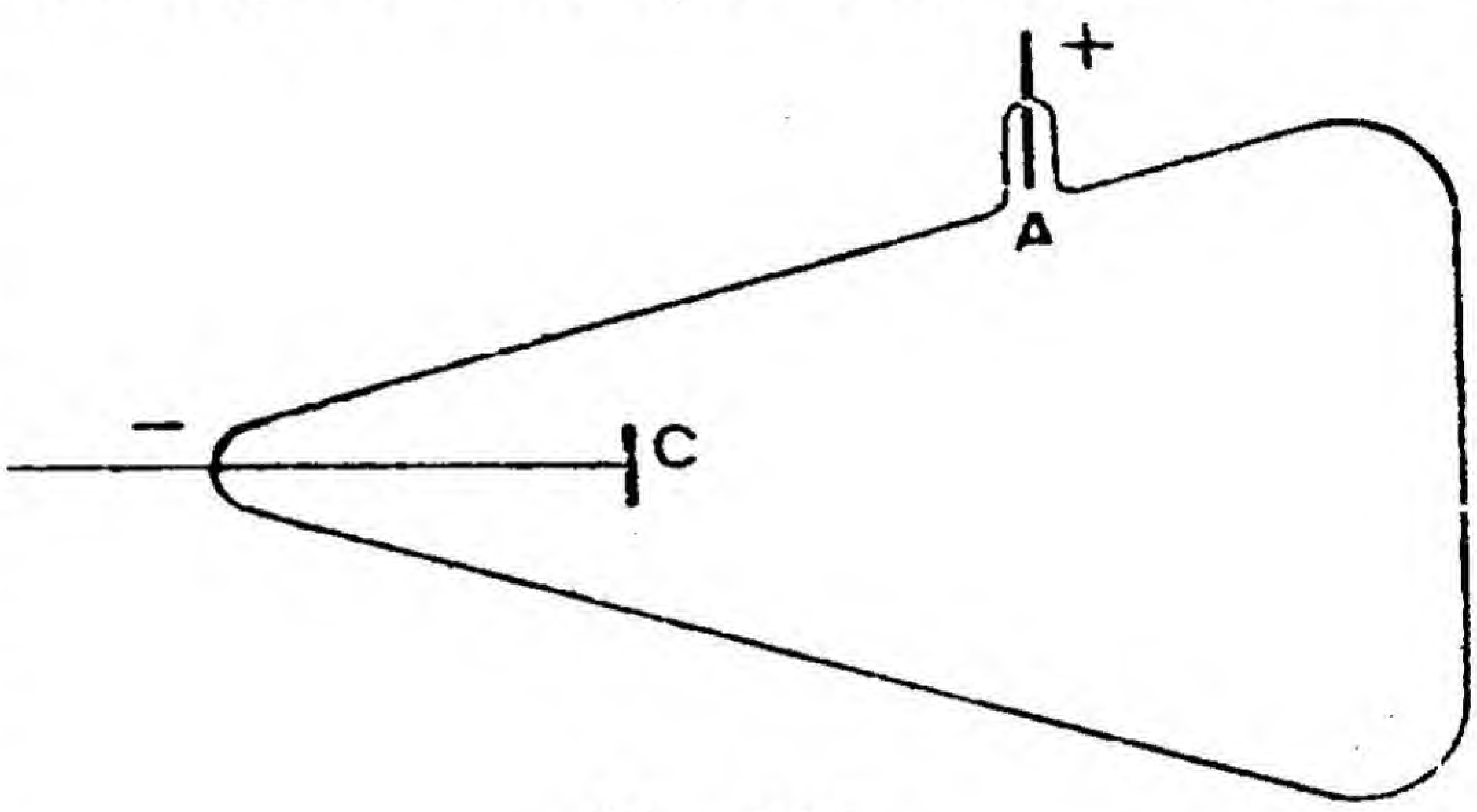


Fig. 265.

backwards, they should have met at the point where the X-rays originated. The lines met in an area rather than a point, and it was about at the place where the cathode rays struck the opposite wall of the tube (Fig. 265). A little later it was shown that the cathode rays were particles of negative electricity travelling at speeds of 2000 miles per second, in fact, they were electrons. **Therefore X-rays are produced when electrons strike a metal target.** It is interesting to notice, in passing, that other physicists said, after the publication of Röntgen's work, that they had noticed that photographic plates near to discharge tubes got fogged, but they had just moved their plates further away. They saved their plates, but they lost the X-rays! Nature whispered to them as she did to Röntgen, but they were insensitive to this hint and they missed a great discovery. This fact should dispose once and for all of the popular fallacy that the discovery of X-rays was an accident! The people who are so fond of saying this could have the same accident happen to them every day of their lives without discovering anything!

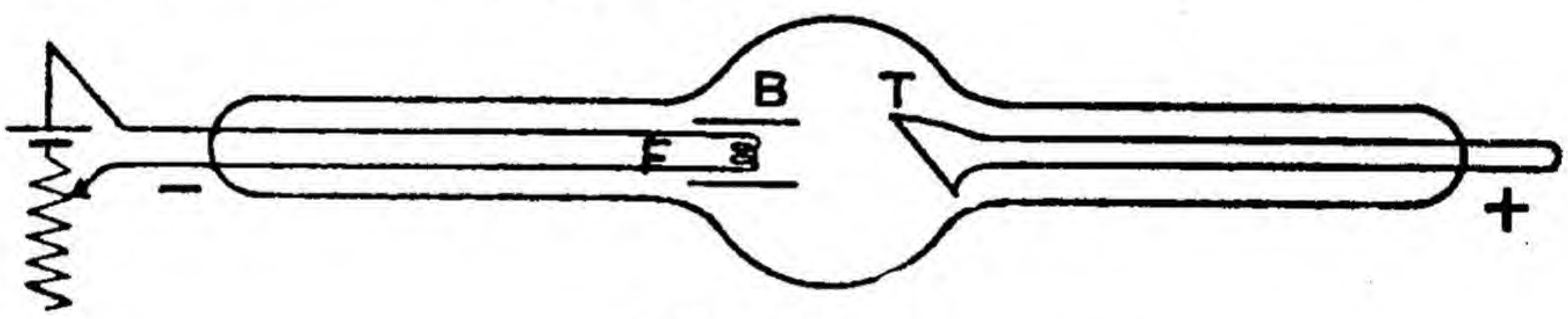


Fig. 266.

The type of tube used by Röntgen was developed into the gas-filled X-ray tube, the term gas-filled referring to the fact that the remnant of gas in the tube is the source of the electrons from which the X-rays are produced. This tube could not be controlled properly as the pressure decreased as it was worked and finally the discharge ceased altogether. It has been replaced by the **Coolidge tube**, which is illustrated in Fig. 266, in which all the factors affecting the nature of the X-rays can be controlled. It consists of a tube, which is exhausted of gas as completely as possible, fitted with a filament F, which can be electrically heated, and a tungsten target T, which also serves as the anode. The filament is connected to the negative terminal and the target to the positive terminal of a high-tension supply of the order of 50,000 volts. When the filament is cold, no current passes as the pressure in the tube is too low. The filament is then heated by an electric current of about an ampere and gives off electrons, which are attracted to the target by the P.D. between it and the filament. A molybdenum tube B round the filament serves to concentrate the electrons on to the target as much as possible. The X-rays are produced by the electrons from the heated filament striking the tungsten target and the properties of the beam are under precise control. If the beam is to be made more intense, it is merely necessary to increase the filament current, when more electrons are emitted from the filament per second and so more strike the target per second, producing a more intense, or "brighter", beam of X-rays. If a more penetrating, or harder,

beam is required, the P.D. across the tube is increased. The beam produced is found to be more penetrating.

The properties of X-rays make them useful, not only in medicine and industry, but also in pure science. They affect a photographic plate, produce fluorescence in suitable minerals, such as barium platinocyanide, and make the air through which they pass a conductor of electricity. They travel in straight lines, they are not deflected by electric and magnetic fields, they cannot be reflected or refracted. **They are absorbed by matter in proportion to its density**; a given thickness of lead absorbs them more than the same thickness of wood; bone absorbs them better than flesh. This differential absorption and the fact of rectilinear propagation are the keys to their use in medicine and industry. If a person's hand is placed between an X-ray tube and a fluorescent screen, a shadow of the bones of his hand is cast on the screen since the flesh hardly absorbs the X-rays at all. Consequently the presence of fractures in bones can be detected and a permanent record can be taken by replacing the fluorescent screen with a photographic plate. If a child swallows a needle, or a soldier is wounded by a bullet, the exact position of the needle or bullet can be found by examination under X-rays. This technique of **shadow photography** can be extended to the organs of the body by filling them with a dense liquid and taking a shadow photograph of the organ, which will be outlined by the dense liquid filling it. Any abnormalities in this outline will help in the diagnosis of the disease. A liquid containing a compound of bismuth or barium is used in the case of the stomach; it is quite immaterial that the bismuth is present in the form of compound, since compounds of bismuth absorb the rays just as strongly as bismuth in the free state. An outline of the complicated system of tubes in the lungs can be made by filling them with lipiodol, a liquid containing a compound of iodine. X-rays are also used in the fitting of shoes, the detection of flaws in the welds of two metal rods, in the examination of aeroplane struts for cracks, and so on.

The use of X-rays in pure science has arisen out of the attempts to find the nature of X-rays. The fact that they are not deflected by electric or magnetic fields rules out the possibility of charged particles and suggests some form of **radiation**, but this view is opposed by the absence of any regular reflection or refraction. If X-rays are another form of radiation, they must be very short waves, partly because there is no room for any other long waves in the spectrum and partly to explain the absence of reflection and refraction. Another possibility put forward was that X-rays are a **neutral doublet**, a close association of a positive and negative charge, travelling at high speed. This would account for all the above facts. Further evidence is needed to settle between these two views. The obvious thing is to look for interference, diffraction, and polarisation. In 1899 Haga and Wind took photographs of the diffraction pattern produced by a beam of X-rays passing through a V-shaped slit and the

result seemed similar to the pattern obtained with light, the bands being obtained only at the narrow end of the slit. This was just what would have been expected if X-rays were very short waves, but the measurements were difficult and the results were not regarded as quite reliable. At any rate no one was convinced enough of the importance of the work to try to repeat the experiment and Haga and Wind's results were forgotten. In 1906 Barkla showed the existence of polarisation in X-rays when he produced a polarised beam by scattering in just the same way that light can be polarised by scattering (Art. 149). This brought fresh support for the view that X-rays were a form of radiation, and in 1912 Laue conceived the idea that *the regular arrangement of the ultimate particles of a crystal would make it an admirable diffraction grating for short waves*, since the distance apart of the atoms was of the order of 1 Å.U., which would enable it to diffract waves as short as this. Let us follow up this idea and see where it leads.

169. THE CRYSTAL AS A DIFFRACTION GRATING

Let us consider a simple crystal such as rock-salt, which crystallises in cubes. We do not yet know the structure of the crystal and we do not even know whether the particles, of which the crystal is built up, are molecules of sodium chloride, or atoms of sodium and chlorine, or some other alternative. So we will refer to them as the particles. Since the crystal is a cube, the particles must be arranged in a cubic fashion, and there will be planes parallel to one face of the cube which are rich in particles. There will be similar planes rich in particles parallel to the other two faces of the cube. Let the dots in Fig. 267 represent one row of particles in one of these planes which is perpendicular to the plane of the paper. If a parallel beam of X-rays falls on this plane, each particle will scatter some of the incident beam; it will act as a point source of X-rays. In fact, each particle will be the source of a weak spherical wave, which we can regard as similar to Huygens' secondary wavelets, and these wavelets will combine to give a reflected wave front making the same angle with the plane as the incident wave front. The plane will therefore give rise to a reflected beam making the same angle with it as the incident beam and of very feeble intensity. The feeble intensity of the reflected beam is due to the number of points at which reflected

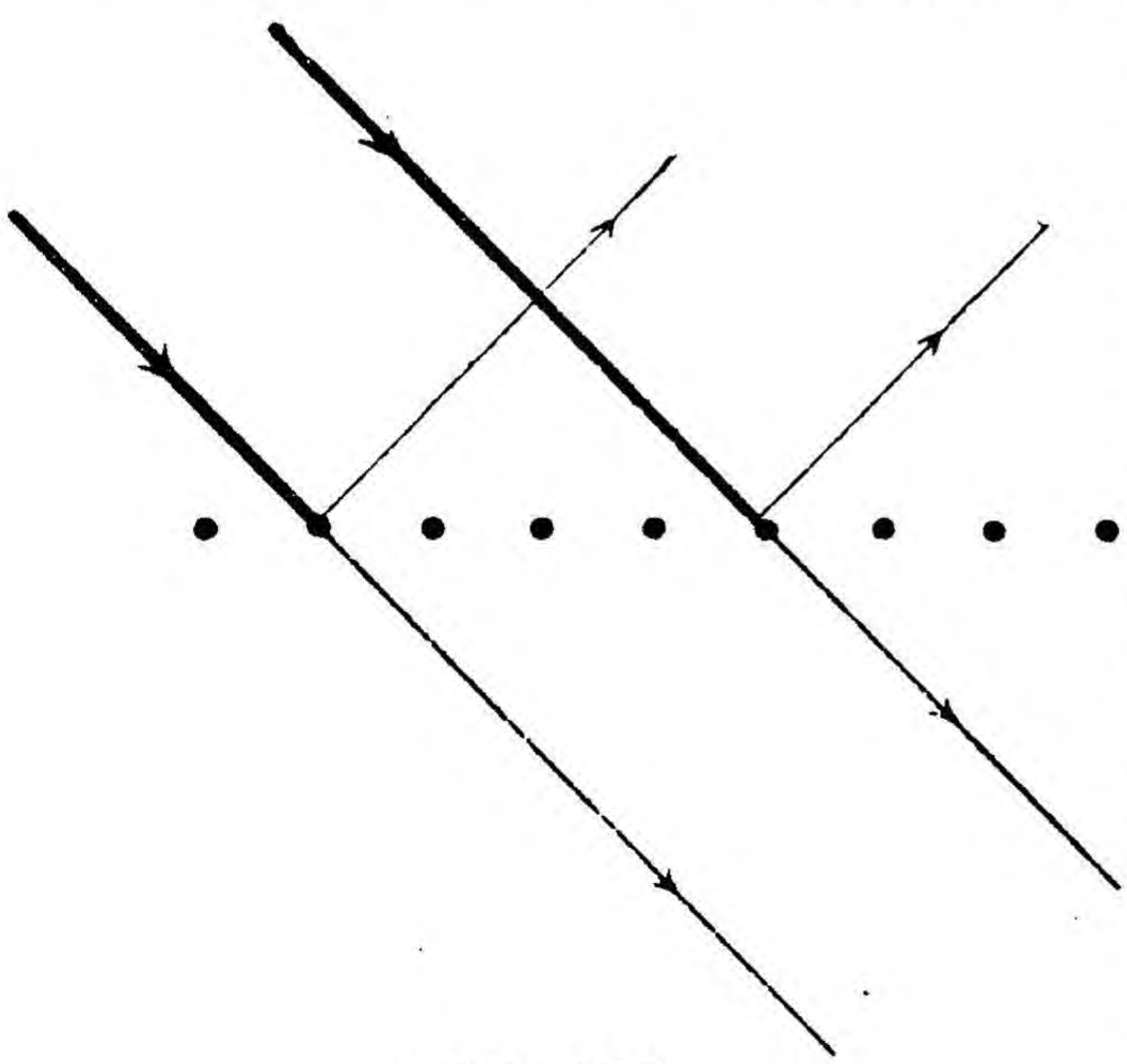


Fig. 267.

wavelets are produced being small compared with the total possible number of such points. The plane of particles, in fact, acts like a plain piece of glass would do for light, only with much less reflecting power. There-

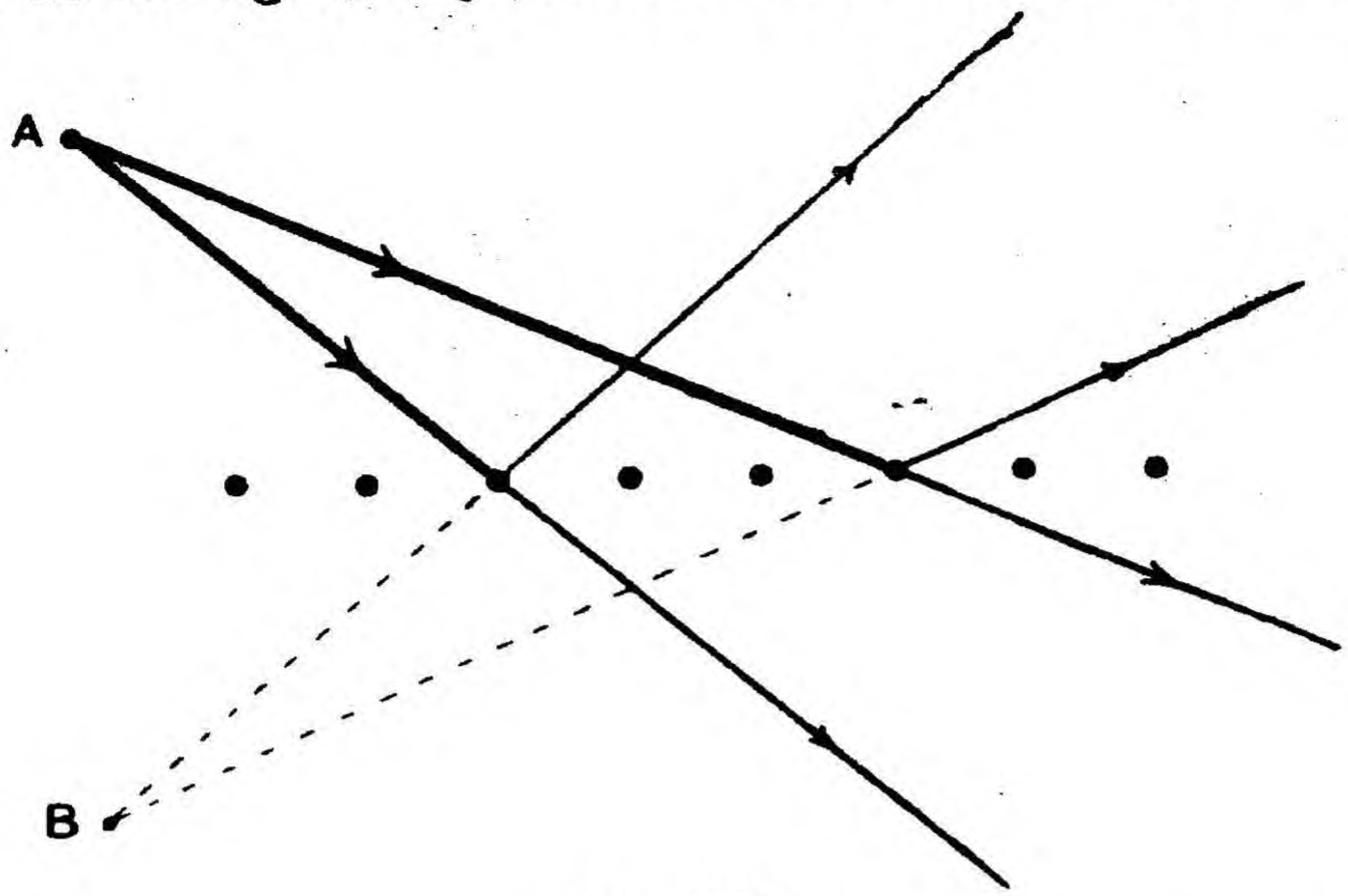


Fig. 268.

fore, if a point source of X-rays A is placed in front of the plane, the plane will produce a virtual image B as far behind the plane as A is in front (Fig. 268). But the reflected ray from this image will never be detected

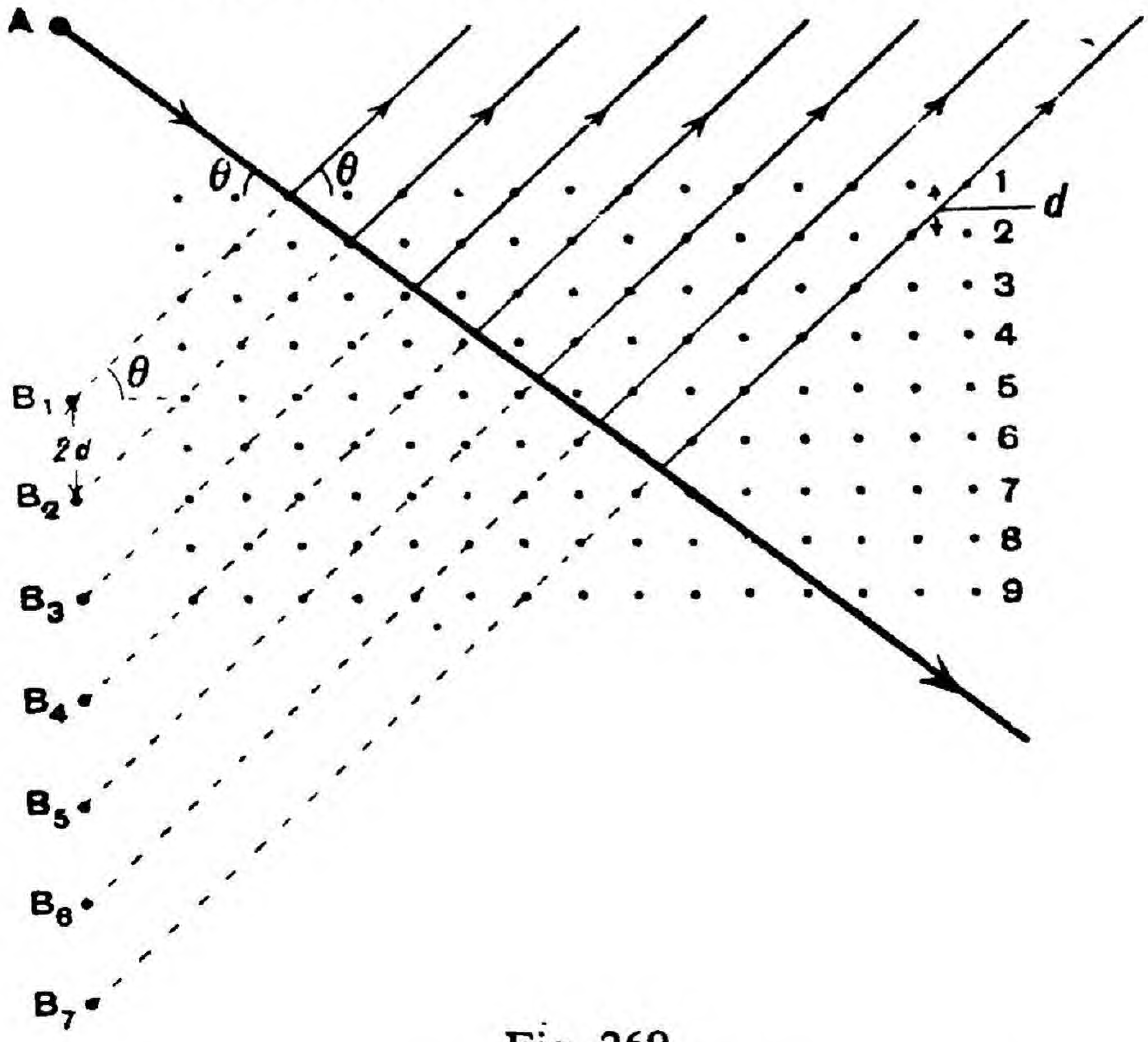


Fig. 269.

by itself, as its intensity is too small. The crystal, however, does not consist of one plane of particles, but of a whole set of planes parallel to and equidistant from one another, which are like the parallel equidistant lines of a diffraction grating. Therefore, if a point source A (Fig. 269)

is placed in front of a crystal, the crystal will produce a set of faint virtual images B_1, B_2, B_3, B_4 . . . by reflection in the successive planes of particles 1, 2, 3, 4, . . . respectively. If we consider one ray from the point source A incident on the crystal at an angle θ, the crystal will produce the reflected rays shown. There will be a reflected beam of maximum intensity in this direction, provided that the individual beams reinforce one another by interference. If the beams are regarded as coming from their virtual sources B_1, B_2, B_3, B_4 . . ., we see that the problem is simply that of the diffraction grating, each source taking the place of a slit or line in the grating. If the planes of particles are a distance d apart, the lines of the equivalent grating are a distance $2d$ apart, so that there will be a beam of maximum intensity in those directions given by the equation

$$2d \sin \theta = n\lambda$$

where λ is the wave-length of the X-rays and n is an integer. It is to be emphasised that θ is the angle between the incident ray and the surface of the crystal and is called the **glancing angle of incidence.**

There are two unknowns in this equation, d and λ; we cannot find both of them. Are we to be thwarted so close to our goal? At any rate, we can accomplish something; we can send a beam of X-rays on to a crystal of rock-salt and see if the reflected beam has a maximum intensity for certain special angles. This turns out to be true, and moreover the ratio of the sines of those angles is as 1 : 2 : 3, as it should be from the above equation. We cannot go into the way in which Sir W. H. Bragg and his son, Professor W. L. Bragg, used the results about the reflected rays to elucidate the structure of rock-salt, **to show that atoms of chlorine and sodium form its particles** and not molecules of sodium chloride, and to find the value of d from the known structure of the crystal and the absolute atomic weights of sodium and chlorine. But d has been found for rock-salt and measuring the values of θ, λ can be calculated. It turns out to be of the order of 1 Å.U., confirming the view of those who contended that X-rays are radiation of very short wave-length. We may mention two final points. Firstly, X-rays have now been reflected and refracted and, secondly, they are proving a powerful tool to investigate the structure of crystals, a matter of immense scientific interest.

170. THE COMPLETION OF THE SPECTRUM

We have now another gap in the spectrum between the ultra-violet, which has been traced down to 1000 Å.U., and X-rays, which go up to a few Angstrom units. This gap has been narrowed down from both sides, the limit of the ultra-violet having been extended by working with the source and a special reflection grating in the same exhausted vessel. In this way radiation of 100 Å.U. has been detected and measured. At

the same time X-rays of longer wave-length have been produced and their wave-length has been measured, not only with crystal gratings, but also with ruled gratings. The gap has been completely closed!

Lastly, the spectrum has been extended to wave-lengths of 0·01 Å.U. both by the discovery of γ-**rays** emitted by radio-active elements and by the production of X-rays of shorter wave-length by designing tubes which will stand higher potential differences. For it can be shown that the greater the P.D. through which the electron falls, the shorter the wave-length of the X-rays produced and the greater their penetrating power. The γ-rays were discovered some thirty years before X-rays of the same wave-length could be produced artificially; they could not be deflected by electric and magnetic fields and they produced a feeble photographic effect and feeble fluorescence. They could penetrate several inches of lead, and so they show a marked similarity to X-rays. This was finally confirmed when their wave-length was measured by the use of a crystal as a diffraction grating. It varies from 1 Å.U. to 0·01 Å.U. Both γ-rays and X-rays are now being used in the treatment of cancer and other malignant growths.

171. SERIES IN SPECTRA

We have now an amazing range of wave-lengths at our command and it is natural to turn to the types of spectra emitted by various substances. We have already seen (Art. 40) that line spectra are produced by atoms and band spectra by molecules, and we can reasonably expect each class to tell us something about the body producing them. We should expect to learn something about the secrets of the atom from line spectra and about the molecule from band spectra. We shall only consider line spectra here and we shall start by saying that, if the complete spectrum of an element such as iron is examined, including lines in the infra-red and ultra-violet, it is seen to be complicated and to show little sign of any regularity. A similar impression is gained by glancing at the spectra of other elements, such as helium or neon. The clue to all this complication and confusion is to be found by examining the spectrum of the simplest element known, namely hydrogen. It consists of five groups of lines, one of which is confined almost entirely to the visible and is shown in Plate VI. This was the first series to be discovered and is known as the **Balmer series**. There is one line in the red, one in the blue-green, one in the blue, one in the violet, the lines getting closer and closer together until they reach a limit just beyond the violet. These lines obviously have some connection with one another. Can we express it in mathematical terms? The really important step was made when the lines were expressed in terms of their frequencies rather than their wave-lengths. All the lines of the series can then be represented by the single equation

$$\nu = R\left(\frac{1}{2^2} - \frac{1}{n_2^2}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (82)$$

where $n_2=3, 4, 5, 6, \ldots$ and R is a constant, called the Rydberg constant. For example, the line in the red is given by $n_2=3$, the one in the blue-green by $n_2=4$, and so on. When the spectrum was extended to the ultra-violet and infra-red, further series were discovered. The **Paschen series** in the infra-red can be represented by the equation

$$\nu = R\left(\frac{1}{3^2} - \frac{1}{n_2^2}\right)$$

where n_2 takes the values 4, 5, 6 . . . The **Lyman series** in the ultra-violet can be represented by the equation

$$\nu = R\left(\frac{1}{1^2} - \frac{1}{n_2^2}\right)$$

where n_2 takes the values 2, 3, 4, 5 . . . Two other series are represented by the equations

$$\nu = R\left(\frac{1}{4^2} - \frac{1}{n_2^2}\right)$$

$$\text{and } \nu = R\left(\frac{1}{5^2} - \frac{1}{n_2^2}\right)$$

respectively. To put it in another way, the frequency of any line in the hydrogen spectrum is given by the equation

$$\nu = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

where n_2 must be greater than n_1. This extension of the spectrum beyond the limits of the visible region now brings to light the fundamental similarity between all these series. Any line in any of the series can be represented as the difference between two frequencies; and one set of frequencies embraces all the series. In other words, any line in the spectrum of hydrogen can be represented as the difference between two of the set of frequencies $\frac{R}{n^2}$, where $n=1, 2, 3, 4, 5 \ldots$. These frequencies are called the **terms** of the hydrogen spectrum; they classify all that is known about that spectrum. They should reveal to us something fundamental about the structure of the hydrogen atom.

What about the spectra of other atoms? We cannot do more here than to say that all line spectra have been reduced to the difference between two terms, each element having its own characteristic set of terms. The frequency of these terms is given by a mathematical relation

of a more complicated kind than that given above. This fundamental fact about the spectra of atoms is remarkable for the fact that the terms form a set of *discontinuous* numbers. Does this indicate a corresponding discontinuity in the structure of the atom itself? It is not certain that it does, for a stretched string has a number of possible modes of vibration whose frequencies form a set of discontinuous numbers, but the stretched string is a continuous medium. We must postpone further consideration of the interpretation of the terms until later (Art. 177), bearing in mind the possibility that it may indicate some kind of discontinuity in the atom.

There is one other striking classification concerning X-ray spectra, which was discovered by Moseley and which yields important knowledge about the atom. Each element gives out a characteristic line spectrum in the X-ray region when bombarded by electrons and the spectrum of a given element consists of a number of series, known as the K series, L series, M series, and so on. Furthermore all these series can be represented as the difference between two terms and *the same equation represents the frequency of the terms for all the elements.* The frequency of the terms in X-ray spectra is given by

$$\nu = \frac{R^2(Z-z)^2}{n^2}$$

where n takes the values 1, 2, 3, 4, . . . R is the Rydberg constant again. It is interesting, in passing, to notice that the K series is given by

$$\nu = R^2(Z-z)^2\left(\frac{1}{1^2} - \frac{1}{n_2{}^2}\right)$$

and the L series by

$$\nu = R^2(Z-z)^2\left(\frac{1}{2^2} - \frac{1}{n_2{}^2}\right)$$

so that they correspond to the Lyman and Balmer series respectively. But the striking thing is that Z is a whole number, the same for all the series, and increasing by unity as we go from one element to the next in the periodic classification. It turns out to be 11 for sodium, 12 for magnesium, 13 for aluminium, and so on. It is equal to the number which would be assigned to an element, if the elements were written down in the order of the periodic classification and numbered hydrogen 1, helium 2, lithium 3, and so on. This number is called the **atomic number** of an element to distinguish it from its atomic weight. It is evidently of fundamental importance in the structure of the atom, since the anomalous positions of potassium and argon, tellurium and iodine, which occur when the periodic classification is drawn up by atomic weight, disappear when it is written out by atomic number. Any theory of the structure of the atom must account not only for its atomic weight but also for its atomic number, and it would seem as if the atomic number were the more

important. We see, then, how the study of spectra leads to some knowledge of the new world inside the atom, which is beginning to open itself before our eyes; we shall pursue these topics a little further in the next chapter by trying to account for spectral terms and for some new facts about radiation which have recently come to light. We conclude this chapter by showing that spectral lines are not only a powerful tool for probing the infinitesimal, but they can also be used to plumb the very depths of space, to explore the great universe of which our home, the earth, is such an insignificant member!

172. THE DOPPLER PRINCIPLE

If we stand on the platform of a railway station and hear an express train whistling as it approaches us, we shall notice a sudden drop in pitch of the whistle at the moment at which the train passes us. The uninformed observer is apt to attribute this drop to the shutting of the tap which admits steam to work the whistle. But this explanation can easily be shown to be wrong in the following ways. If a number of observers are stationed at intervals along the platform they will all report a drop in pitch at the precise moment at which the engine passed them. The driver cannot close the tap as many different times as there are different observers! Secondly, a similar effect can be noticed when the train passes without whistling at all. In this case, the rate at which the puffs of steam come out of the funnel suddenly decreases at the moment at which the train passes the observer. Both of these facts are particular cases of the Doppler principle, according to which the frequency of any source of waves as recorded by a given observer depends on the velocity of the source relative to the observer. **The frequency of a source of waves recorded by an observer is greater when the source is moving towards the observer than when it is at rest relative to him and it is less when the source is moving away from the observer than when it is at rest relative to him.** Accordingly the observer on the platform hears a higher pitch from the whistle than the driver on the footplate of the engine, while the engine is approaching, and a lower pitch when it is receding. It must be emphasised that there is no question of the pitch recorded by the observer on the platform changing as the train gets nearer; the pitch remains constant all the time during which the train approaches and the drop is confined to the short interval of time during which the approach changes to recession.

The relation between the frequency of a source of waves recorded by an observer and his velocity relative to the source can be investigated mathematically in the following way. Let u be the velocity of a source of waves S (Fig. 270), let v be the velocity of an observer O, let c be the velocity of the waves relative to the medium, and let w be the velocity of the medium itself. All these velocities count as positive when they

are to the right and u, v, and w are measured relative to some convenient body of reference, such as the earth. Let n be the frequency of the waves emitted by the source. The reader may at once ask: the frequency measured by what observer? We will not specify an observer for the moment; we will suppose that the frequency of the source of waves can be measured without making observations on the waves themselves. For example, this is quite possible in the case of a tuning fork; we can measure the frequency of vibration of the fork itself. In the cases of other types of wave we can *imagine* ourselves doing a similar sort of operation. We have the general equation

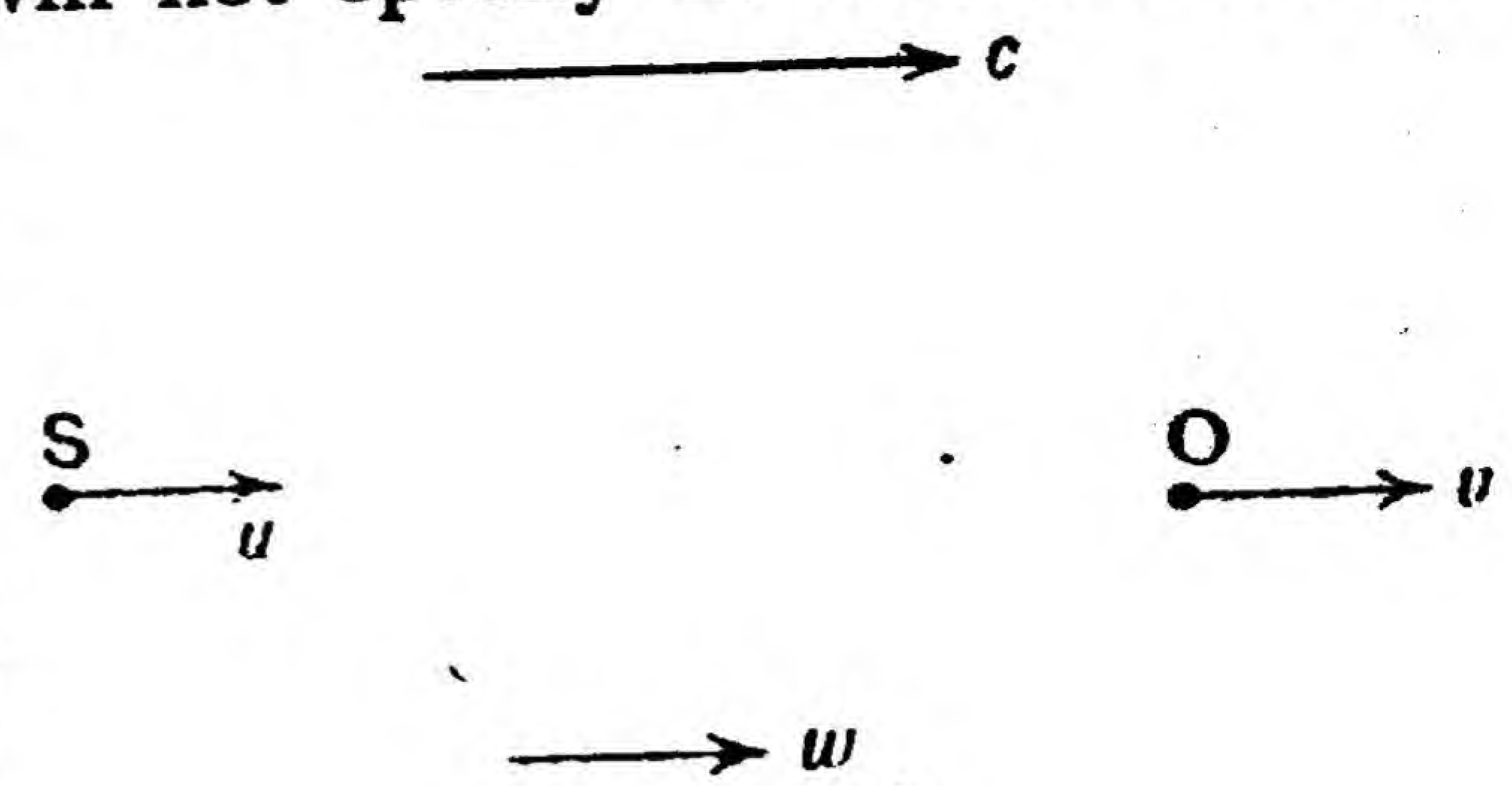


Fig. 270.

$$c = n\lambda$$

or

$$\lambda = \frac{c}{n}$$

In general, when the source and the medium are moving, this can be written:

$$\text{Apparent wave-length} = \frac{\text{Velocity of waves relative to the source}}{\text{Frequency of the source}}$$

In the present case, if λ' denotes the apparent wave-length

$$\lambda' = \frac{c+w-u}{n} \quad \text{. . . . . (83)}$$

We also have

$$n = \frac{c}{\lambda}$$

which becomes, in the case when the medium and the observer are moving,

$$\text{Frequency recorded by observer} = \frac{\text{Velocity of waves relative to observer}}{\text{Apparent wave-length}}$$

So, in this case, if n' is the frequency recorded by the observer,

$$n' = \frac{c+w-v}{\lambda'}$$

Substituting the value of λ' from equation (83) in this equation, we have

$$\frac{n'}{n} = \frac{c+w-v}{c+w-u} \quad \text{. . . . . . . (84)}$$

This equation shows us at once the meaning of n; for $n=n'$, when $u=v$, and so **the frequency of the source is just its frequency recorded by an observer at rest relative to it.** We also see that, if the source approaches the observer, $u>v$ and so $n'>n$, while, if it recedes from the observer, $u<v$ and $n'<n$. These deductions agree with the observations on a train going through a station whistling.

The equation can be applied most simply to cases in light by putting $w=0$ and working in wave-lengths, λ and λ', corresponding to the frequencies n and n' respectively. Treating the observer as at rest, equation (84) then becomes

$$\frac{\lambda'}{\lambda}=\frac{c-u}{c} \qquad (85)$$

Therefore the wave-length of a source of monochromatic light will be shifted towards the violet if the source is approaching and towards the red if it is receding, and, from the amount of shift, the velocity of the source can be calculated. For example, if the source is approaching at 10 m.p.s., which is about the velocity of the earth in its orbit round the sun,

$$\frac{\lambda'-\lambda}{\lambda}=-\frac{u}{c}$$

$$\therefore \frac{\lambda'-\lambda}{\lambda}=-\frac{10}{186{,}000}$$

If

$$\lambda=6000 \text{ Å.U.},$$

$$\therefore \lambda'-\lambda=-\tfrac{1}{3} \text{ Å.U. approximately.}$$

The difference in wave-length of the sodium D lines is 6 Å.U., so that the shift can be detected in the case of velocities of the order of miles per second with a spectrometer of high resolving power.

173. APPLICATIONS OF THE DOPPLER PRINCIPLE

This result, which has been verified in the laboratory for light waves as well as being well known in the case of sound, has a number of interesting applications in astronomy. The presence of sunspots enables us to deduce that the sun is rotating on its axis about once in 27 days, since sunspots are observed to move across the sun's disc at a rate compatible with this period of rotation. The rotation can be confirmed by pointing the slit of a spectrometer at the east limb of the sun and photographing the Fraunhofer lines; a similar photograph is taken with the slit pointing at the west limb of the sun. The corresponding lines are found to be slightly displaced relative to one another, those corresponding to the west limb being displaced to the violet relative to those from the east limb. This indicates that the sun is rotating from west to east and, from

the amount of the displacement, the velocity of the edge of the sun can be calculated. It agrees with the above period of rotation.

The nature of Saturn's rings was also deduced from Doppler's principle. The rings may be either solid discs or a swarm of moons. In the first case, the outer edge will move faster than the inner, while, in the second case, the precise opposite follows from the law of gravitation. The matter can be settled by pointing the slit of a spectrometer first at the inner edge and then at the outer edge of Saturn's rings and taking photographs of the Fraunhofer lines in each case. It is found that the lines show more displacement from the normal position in the case of the inner edge than for the outer edge, thus proving that Saturn's rings are a swarm of moons.

The velocity of stars in the line of sight can also be measured by pointing a spectrometer at the star and photographing its spectrum. This contains some absorption lines similar to the Fraunhofer lines, and their general position in the spectrum enables the element producing them to be identified, which fixes their wave-length for an observer at rest relative to the source. When the wave-length of the line in the spectrum of the star has been measured, the velocity in the line of sight can be calculated. The lines are displaced to the violet if the star is approaching and to the red if it is receding. Stars moving with velocities of several miles per second have been detected in this way.

It is known that a number of stars are describing orbits round one another, for they can be observed doing so in a telescope. Such stars are called binaries. But some stars have been observed which appear single in a telescope and whose spectra show a periodic variation. The lines are now single, then each line splits into two lines after a certain time, after an equal time they are single again, and so on. This is due to the fact that the star is a binary, with the two components too close together to be resolved in the telescope. When the lines are single, each component is moving at right angles to the line joining them to the earth and the spectral lines are in their normal positions ; when the lines are double, one star is approaching and the other receding from the earth, the lines of the former being displaced to the violet, while those of the latter are displaced to the red. Such stars are called spectroscopic binaries, and many single stars have turned out to be spectroscopic binaries.

Perhaps the most startling discovery by Doppler's principle has been made in the last five years, when it has become possible to measure the shift of the spectral lines in the spiral nebulæ. These are island universes of stars similar to the collection of stars of which the sun is one, but lying outside our system. The spectral lines of these nebulæ are all displaced to the red, the displacement increasing the more distant the nebula. The velocities found are very big, some of the nebulæ going at 800 km. per sec. A nebula in the neighbourhood of the Gemini at a distance of 150,000,000 light years is receding at a velocity of 15,000 miles per sec. !

If these undoubted shifts of the spectral lines to the red are accepted as being caused by the recession of the nebulæ, then the conclusion of all these measurements is that the universe is expanding! Some workers have alternative explanations of the red shift which does not involve a recession of the nebulæ, but we have evidently detected a phenomenon of fundamental importance for the interpretation of the universe.

The seed which Newton sowed when he commenced his experiments on white light in order to improve the images of telescopes has borne good fruit! Newton replaced the colour discrimination of the eye as a means of distinguishing between different sorts of light by a physical quantity, refractive index. This has been superseded in its turn by wave-length. While still seeking a solution of Newton's original problem, Fraunhofer discovered the dark lines in the sun's spectrum, which led to the discovery of line spectra. Then the spectrum was extended in both directions, until the original visible spectrum has shrunk into insignificance and the conception of light itself has been enlarged to embrace a wave motion in free space whose wave-length may vary from atomic to terrestrial magnitudes. Lastly the spectral lines have proved to be a tool for probing the vast extent of the universe in which we live by analysing the radiation which is sent to us from its remotest depths, and for unravelling the mysteries which lie hidden within that new microscopic universe inside the atom itself.

EXAMPLES ON CHAPTER XVI

1. Compare the glass prism and plane diffraction grating as instruments for the analysis of spectra in the optical region. Explain why different materials are used for the construction of prisms for use with infra-red radiation, and indicate the type of modification you would expect to find in gratings which were to be similarly used. *(Camb. Schol.)*

2. Describe how infra-red radiation may be obtained and how its properties may be studied. How may the wave-length of a monochromatic beam of infra-red radiation be determined? *(Camb. Schol.)*

3. What reasons have we for supposing that X-rays, visible light, and ultra-violet and infra-red radiation are essentially similar in nature? Compare and contrast briefly their more important properties. *(Camb. Schol.)*

4. What are the factors which affect the diminution in intensity which a parallel beam of white light suffers when passing through the air? What advantages are gained by using photographic plates sensitive to light in the infra-red only? What would be the characteristics of photographs taken with plates sensitive only to ultra-violet light or of those taken in a thin fog? *(Camb. Schol.)*

5. Give an account of the methods which can be used in the investigation of the infra-red and ultra-violet regions of the spectrum.

Enumerate the types of electro-magnetic radiation other than visible light, giving their sequence in order of wave-length. *(Camb. Schol.)*

6. A parallel beam of radiation falls on a 60° quartz prism, which is set so that the wave-length 12,560 Å.U. of refractive index 1·532 passes through the prism at minimum deviation. The emergent radiation is focussed on a screen by a

concave mirror of focal length 50·0 cm. Find the length of the infra-red spectrum between wave-lengths 7500 Å.U., refractive index 1·540, and 42,000 Å.U., refractive index 1·457; find also the length of the visible spectrum between 7500 Å.U. and 4000 Å.U., refractive index 1·557. Ignore any double refraction.

7. A source of radiation rich in visible and ultra-violet gives out a line in the red, one in the ultra-violet, and one in the far ultra-violet, whose refractive indices in quartz are 1·540, 1·577, and 1·631 respectively. The source is placed 20·0 cm. from a converging lens whose focal length is 20·0 cm. for the line in the ultra-violet. Another converging lens, focal length 30·0 cm. for the same line, is placed 10·0 cm. beyond the first lens. Find the distance from the second lens at which the three lines come to a focus ignoring double refraction. Use your results to explain why the plate of an ultra-violet spectroscope is tilted at about 20° to the axis of the lens focussing the spectrum.

8. A concave reflection grating of radius of curvature 3 metres and 5000 lines per cm. is used with Rowland's mounting, in which S is the source and P denotes a line in the spectrum on the photographic plate (Fig. 259.) Find the distance SP for the following iron arc lines in the first order spectrum, 4022 Å.U., 3640 Å. U., 3030 Å.U., and 2813 Å.U.

9. How would you measure the wave-length of the D lines of sodium? The light from the star Sirius contains the blue-green hydrogen line (F) whose wave-length is found to be 4862·12 Å.U. When this line is produced in the laboratory its wave-length is found to be 4861·37 Å.U. What deductions can you make from these figures? (Velocity of light $=3\times10^{10}$ cm. per second.) *(Oxford Schol.)*

10. The wave-length of a spectrum line coming from a star is changed by the motion of the star from 6000 Å.U. to 6001 Å.U. Find the velocity of the star relative to the earth. (Velocity of light $=3\times10^{10}$ cm./sec.) *(Camb Schol.)*

11. What effect is produced on the apparent wave-length of light waves
(a) when the source moves and the observer is at rest;
(b) when the source is at rest and the observer moves?
Discuss some applications of this effect to astronomy, and to the broadening of lines emitted by a hot gas. *(Camb. Schol.)*

12. What is the Doppler effect and how can the velocities of the stars be measured its use? *(Camb. Schol.)*

13 Explain Doppler's principle.
A spectrum line of wave-length 4×10^{-5} cm., in the spectrum of the light from a star, is found to be displaced from its normal position towards the red end of the spectrum by an amount equivalent to 10^{-8} cm. What velocity of the star in the line of sight would account for this? *(Tripos, Part I.)*

14. If a photograph of the sun's spectrum is taken with the slit of the spectrograph pointing to (a) the eastern edge, (b) the western edge of the sun at its equator, corresponding lines in the two spectra are displaced a little relative to one another. Explain this and calculate the displacement in the case of the F line, taking the diameter of the sun as $1{\cdot}4\times10^{6}$ km. and the period of rotation at the equator as 25 days. The normal wave-length of the F line as measured by apparatus at rest relative to the source is 4861 Å.U. An elementary physics laboratory will usually have a diffraction grating of about 5500 lines per cm. and 3 cm. long. Would it be possible to detect the above displacement with such a grating?

15. Atoms of hydrogen moving with velocities of 10^{8} cm. per sec. are produced in a discharge tube. Calculate the wave-length which the C line of such a hydrogen atom would have if it was moving directly (a) towards, (b) away from the slit of the spectrometer. The wave-length when it is at rest relative to the spectrometer is 6536 Å.U.

16. Give details of the equipment required for the production of a beam of X-rays.
Explain the general principle of a method for determining the wave-length of a beam of homogeneous X-rays. *(Tripos, Part I.)*

17. Show how a crystal can act as a diffraction grating for radiation of wave-length of the order of 10^{-8} cm. and derive an equation relating the angles of diffraction and the wave-length of the radiation. If the distance between successive planes rich in atoms in rock-salt is $2{\cdot}80 \times 10^{-8}$ cm. and the glancing angle in the first order spectrum for a particular X-ray line is $11{\cdot}4°$, calculate the wave-length of the X-rays. Find also the glancing angle for this wave-length in the second and third order spectra.

Chapter XVII

THE RENAISSANCE OF THE CORPUSCULAR THEORY

174. THE PHOTO-ELECTRIC EFFECT

We now seem to have come to an end of our enquiry as to the nature of light, for we have consistent and overwhelming evidence in favour of the view that light is transverse waves in a subtle medium, the ether, pervading the whole of space. There is still the question as to the precise nature of the condition which is propagated in light waves and bound up with this is the nature of the ether itself. But we have already said that we cannot pursue this topic further in this book, frankly admitting that the idea of transverse waves in a medium more rare than air itself is a most difficult one and that some solution of this problem must be sought. But just at the moment when the wave theory was at the height of its triumph one or two new phenomena were discovered demanding attention. In 1888 Hallwachs discovered that when a zinc plate is exposed to ultra-violet light it acquires a positive potential of the order of 1 volt. This fact is known as the photo-electric effect, for it concerns the conversion of light into electricity. Some years later it was discovered that the positive potential was caused by the emission of electrons by the plate, the electron being the atom or unit of negative electricity. The question naturally arises: how does the number and velocity of the emitted electrons depend on the character of the incident radiation, that is, on its intensity and frequency? What answer, based on the wave theory, can we make to these questions?

It is difficult to give a precise answer without a clear mechanism of the way in which the light causes the emission of the electron, which, in turn, demands some knowledge of the structure of matter. The reader will be familiar with the fact that an element consists of atoms separated by distances about equal to their own diameters, which are of the order of 10^{-8} cm. Each atom is to be regarded as a miniature solar system, consisting of a minute positive nucleus with some planetary electrons describing orbits round it. The nucleus consists of protons, the unit of positive electricity, and electrons, the positive charge being due to excess of protons. The number of planetary electrons is just sufficient to make the atom electrically neutral. The outer planetary electrons are attracted

only feebly by the positively charged nucleus and are the ones which will be ejected from the metal by the radiation falling on it. Since the radiation only penetrates a layer about 10^{-7} cm. thick at the surface of the metal, it is just the outer planetary electrons of the atoms near the surface which are emitted. When the light wave falls on the metal, some of the electrons absorb enough energy from the wave front to overcome the attraction of their positive nuclei and to be ejected from the metal with a finite velocity, the rest receiving additional energy of vibration while remaining in the atom. This energy is ultimately passed on to the atom as a whole, that is, the metal warms up. If the intensity of the incident radiation is increased, the amplitude of the waves is made bigger in the language of the wave theory. More energy falls on the electrons now, so we should expect them to be emitted with a greater velocity. We should not expect any marked increase in the number of emitted electrons, since the wave front covers the whole of the metal plate in each case and the number of electrons affected by it is likely to be the same in each case. It is true that a wave of greater amplitude will penetrate more deeply before all its energy is absorbed by the metal, but the electrons from the deeper layers would hardly be expected to escape from the metal plate, as the light wave would be of too small an amplitude when it reached them to give them sufficient energy for the purpose. **So we expect an increase in the intensity of the incident radiation to produce an increase in the velocity of the emitted electrons.** To put it in another way, the closer the source of radiation is to the metal the faster the electrons should be emitted. A similar case with water waves may make the argument more convincing. If a huge boulder is dropped into a lake near to a fishing boat fixed in position, the wave produced will hit the boat with considerable violence and may give it such a blow as to hurl a stone lying in the bottom of the boat many feet into the air. The closer the boat is to the point at which the boulder strikes the lake, the faster the stone will leave the boat and the further it will go into the air.

A change in the frequency of the radiation, on the other hand, should produce little effect on the emitted electrons, since the frequency of the waves has far less influence on their energy than the amplitude, especially if we restrict ourselves to the visible spectrum. This is because the energy of a wave is proportional to the square of both the amplitude and the frequency, but, while the amplitude may be increased a hundred times by an increase in intensity, the frequency can only be doubled in going from red to violet. It is possible that there may be some kind of resonance effect, akin to the great increase in volume when a wireless-set is tuned to an incoming wave; in the photo-electric effect, we should expect a specially large velocity of emission round about a frequency characteristic of the metal. But, with this reservation, we may say that the frequency of the incident radiation should have only a secondary effect on the velocity of the emitted electrons.

What are the facts? They are very difficult to establish, since the electrons have so little energy that it can easily be altered by the thin film of gas which is absorbed on the surface of any metal. So the progress was slow in the early stages and different workers obtained different results and the same worker often found it difficult to repeat his results. And a result cannot be accepted as a fact in science until it is common to all the observers (Art. 1). But in 1902 Lenard succeeded in establishing the fact that **the velocity of the emitted electrons is independent of the intensity of the incident radiation.** A little later the still more mysterious fact emerged that no electrons are emitted at all unless the frequency of the incident radiation, assumed monochromatic, is greater than a certain critical value, called the **threshold frequency**, which is different for each metal. This frequency is in the green for sodium, while it is in the ultra-violet at a wave-length of 3×10^{-5} cm. for zinc. How can these two facts be explained on the wave theory? Is the threshold frequency the expected resonance effect? The idea is tempting, but further investigation renders it untenable, since the emission continues with increased velocity when the frequency of the incident radiation is greater than the threshold frequency. But, on a resonance theory, the emission should fall off for frequencies above, as well as for those below, the critical value, just as the volume produced by a wireless-receiver diminishes whether the frequency of the set is less than or greater than that of the incoming wave. And how are we to explain the independence of the intensity of the radiation and the velocity of the emitted electrons? Perhaps a way out is indicated by the following numerical case. If a standard candle is placed 2 metres from a metal plate, it sends light on to it at the rate of 1 erg per sq. cm. per sec. Assuming that one electron can absorb all the light which falls on the cross-section of the atom of which it is a part, the electron absorbs 10^{-15} ergs per sec., taking the area of cross-section of an atom as 10^{-15} sq. cm. The electrons are observed to be emitted with an energy of 10^{-12} ergs. On the wave theory, then, 1000 secs., or more than a quarter of an hour, should elapse after switching on the light before an electron is emitted. As a matter of fact the emission is instantaneous! The only possible way of explaining facts of this kind on the wave theory is to assume that the light acts as a kind of "trigger," which ejects the electron from the atom by releasing some energy stored in the atom for that purpose. On this view we should expect the velocity of the emitted electrons to be independent of the character of the incident radiation, since that velocity is obtained, not from the energy of the incident radiation, but from an independent supply inside the atom.

175. THE QUANTUM THEORY OF RADIATION

The above facts and hypotheses were being discussed in the first few years of the twentieth century, when Einstein published a bold and

revolutionary theory in 1905. The facts of black body radiation, which had defied explanation on the classical mechanics, were successfully explained by Planck in a paper published in 1900. The essential feature of his theory is that the energy of the oscillators in the body emitting the radiation cannot assume any value, but only integral multiples of the **quantum** of energy, $h\nu$, where ν is the frequency of the oscillator and h is Planck's universal constant; its value is $6{\cdot}62\times10^{-27}$ ergs × secs. The body turns some of its energy into radiation when one of these oscillators loses one quantum of its energy, emitting it as radiation of energy $h\nu$ and of frequency ν. In Planck's original form of the theory both the emission *and* absorption of radiation are discontinuous. But Planck saw that a discontinuous absorption implies that radiation is corpuscular in nature, which conflicts with the idea of a spherical wave front spreading out in space, a conception which is essential to explain interference and therefore diffraction. So he modified his theory in order that discontinuous absorption might not be necessary and was still able to account for the facts of black body radiation.

Einstein asserted that Planck's original version of his theory is the true one and that the conception of radiation as composed of bundles of energy travelling on in straight lines as discrete units is the clue needed to unravel the curious and contradictory facts of photo-electric emission. He called the corpuscles, or bundles of energy, **photons** and he pictured the emission of electrons, when radiation falls on a metal plate, in the following way. If the incident radiation is of frequency ν, it sends photons of energy $h\nu$ on to the metal and an electron can absorb the whole of the energy of one photon. But we have already seen that a certain amount of work, w, is needed to extract the electron from the nucleus of its atom and to get it out of the surface of the metal. It follows at once that, if $h\nu$ is *less* than w, there can be no emission however great the intensity of the incident radiation. For an increase in intensity merely increases the *number* of photons falling on unit area of the plate in unit time, but it does not change the amount of energy in each photon. Hence an electron can never get enough energy to bring it out of the surface of the metal. So the threshold effect follows quite naturally on this theory, the threshold frequency ν_0 being given by the equation

$$h\nu_0=w \qquad (86)$$

If the frequency ν of the incident radiation is greater than ν_0, then an amount w of the energy of the photon is used in bringing the electron to the surface of the metal, the remainder being given to the electron in the form of kinetic energy. The electron will accordingly be emitted with velocity v, where

$$\tfrac{1}{2}mv^2=h\nu-w \qquad (87)$$

Combining these two equations, we have

$$\tfrac{1}{2}mv^2=h(\nu-\nu_0) \qquad (88)$$

The quantum theory therefore predicts that **the velocity of the emitted electrons is independent of the intensity of the incident radiation, but increases as the frequency increases** in accordance with equation (88). An increase in intensity of the radiation means an increase in the number of photons striking the plate per second and therefore an increase in the number of electrons emitted in that time. These predictions, so boldly made by Einstein at the time when the experimentalists spoke with an uncertain and divided voice, are the very opposite to what is expected on the classical theory, although they have the merit of fitting the two quantitative facts which were well established. What is to be their fate? They were vigorously attacked by the supporters of the classical wave theory, chiefly on the ground that a corpuscular theory is quite unable to explain interference and diffraction. And there was another objection too: what are we to make of a theory which replaces waves by corpuscles and then specifies the energy of those corpuscles in terms of a frequency, which tacitly implies a vibratory, and therefore undulatory, character for the light? The supporters of the wave theory pointed out that if contradictory assumptions are made, anything can be proved. But Einstein held calmly to his position, ignoring the difficulties for the moment. He felt convinced that his clue is the right one for this particular class of facts about light, which are distinguished from interference and diffraction in that they concern the interchange of energy between matter and radiation. He allowed the facts to dictate his ideas and was not concerned by the impossibility of getting a mechanical model for them. No doubt he hoped that the contradiction between his theory and the wave theory would be settled by a more comprehensive theory embracing both his quantum theory and the wave theory. Was not his attitude the same as that of Newton when faced with the facts of rectilinear propagation and Newton's rings, one seeming to demand a corpuscular theory and the other waves? Newton boldly modified his corpuscular theory to introduce waves and now we have a complete wave theory which does account in a consistent way for both of the above facts.

At all events, the marked difference in the predictions of the wave theory and Einstein's quantum theory stimulated experimental work and the variable retardation of the emitted electrons by thin surface films of gas was overcome by preparing the metal surface in a vacuum, when consistent values for the velocity of the emitted electrons were obtained for the first time. This advance in technique was made in England by Hughes and was later used by Millikan in America. In 1917, Millikan finally proved the truth of Einstein's equation to an accuracy of 1 per cent., his results providing one of the most accurate determinations of Planck's constant which have ever been made. Using the alkali metals prepared in a vacuum, he sent radiation of known frequency ν to the metal surface and measured the maximum velocity v of the emitted

electrons by finding the potential V just needed to prevent any electrons from reaching a metal cylinder near to the plate exposed to the incident radiation. Then

$$Ve=\tfrac{1}{2}mv^2$$

and so

$$Ve=h(\nu-\nu_0)$$

Millikan found that the graph of V against ν is a straight line, thus verifying Einstein's theory. The reader should consult Millikan's "Electron" for full details of this experiment. Einstein's quantum theory gives a complete and consistent explanation of the photo-electric effect without the necessity of any *ad hoc* assumptions. But the wave theory met with difficulties early on, and had to make a new assumption for each new fact which was discovered. This state of affairs sounds the death knell of a theory, and Millikan's results render the trigger hypothesis untenable. For, if the energy of the emitted electron comes from the atom itself, why is it changed by changing the character of the incident radiation? It might be possible to invent some complicated kind of mechanism to produce this result, but it would be little use in suggesting further lines of enquiry, since it would be so complicated. Again, such a mechanism would give little æsthetic satisfaction. Perhaps this æsthetic element is the final court of appeal in deciding between the merits of two scientific theories. It is not usually true to say that a set of facts can be explained by only one theory; there are possible alternatives. But one of the alternatives will make the greatest appeal on account of the simplicity of its fundamental postulates and the beauty of the reasoning leading to the consequences of those postulates. To this extent science is not only an intellectual adventure, an attempt to provide an all-embracing and simple correlation of an ever wider range of facts, it is a form of art itself. So we must abandon the attempt to explain the photo-electric effect on the wave theory and accept a corpuscular theory of radiation.

176. THE COMPTON EFFECT

So far we have seen that radiation reveals a corpuscular nature when it is converted into the kinetic energy of electrons. Now we shall see that it reveals the same nature even when it is preserving its own existence! This has been shown in connection with the scattering of radiation by free electrons. On the classical theory, the scattering happens in some such way as this: the incident radiation of frequency ν sets the electrons into forced vibration with the same frequency. These vibrating electrons react back in turn on the ether and radiate light of their own frequency, that is, a frequency ν. Hence the frequency of the scattered radiation is the same as that of the incident radiation. While measurements were only approximate, this was found to be true and has long been known

to be true for the scattering of light by small particles. On the quantum theory, however, the scattering of radiation by electrons is to be treated as **an impact between the photons and the electrons obeying the laws of the conservation of energy and momentum.** It follows from the theory of relativity that a quantity of energy E has a mass $\frac{E}{c^2}$. Therefore a photon of energy $h\nu$ has a mass $\frac{h\nu}{c^2}$ and a momentum $\frac{h\nu}{c}$. So the problem can be treated as an impact between the photon of mass $\frac{h\nu}{c^2}$ and velocity c and an electron of mass m and no velocity. There are two possible cases; if the electron is bound to the nucleus of its atom, the impact of the photon will cause the atom as a whole to recoil. Now the mass of an atom is of the order of 10^{-24} gm., while that of a photon of X-rays is of the order of 10^{-28} gm. Therefore, by the conservation of momentum, the photon can communicate a negligible velocity, and therefore negligible energy, to the atom. Hence it is scattered with unchanged energy, that is, with unchanged frequency on the quantum theory. But, if the electron is free, if it is not bound to the nucleus of its atom, the impact is one between the photon and the electron itself. The mass of the electron is only some ten times greater than that of a photon of X-rays and so, by the conservation of momentum, the photon can contribute a finite velocity to the electron. **Hence the electron recoils with a finite velocity and energy and, by the conservation of energy, the photon must be scattered with decreased energy and therefore diminished frequency.** The increase in wave-length, $d\lambda$, corresponding to this decrease in frequency can be calculated from the equations expressing the conservation of energy and momentum and is given by

$$d\lambda = \frac{2h}{mc} \sin^2 \tfrac{1}{2}\theta$$

where m is the mass of the electron, c is the velocity of light, and θ is the angle through which the photon is scattered. The change in wave-length for $\theta = 90^\circ$ is $0{\cdot}0242 \times 10^{-8}$ cm. Therefore, if a beam of X-rays is sent on to a metal rich in free electrons and the spectrum of the scattered X-rays is photographed, there should be a line of the same frequency as the incident beam due to scattering by bound electrons and one of wave-length $0{\cdot}0242 \times 10^{-8}$ cm. greater just beside it.

Compton claimed to have found this effect in 1924, using a carbon scatterer, since carbon is rich in free electrons. When Duane and Clark sought for the effect using the same material, they could not find any variation of the change in wave-length with the angle through which

the light is scattered and they suggested an alternative explanation of the effect observed by Compton. Finally Watson, Becker, and Smyth at Pasadena in California found both the line of unchanged wave-length and the one of greater wave-length using aluminium as the scattering element. They found the required variation of change in wave-length with scattering angle, their results being accurate to 1 per cent. This Compton effect, as it is called, is perhaps the most vivid and convincing piece of evidence in favour of the view that light is corpuscular in nature. There is no escaping this conclusion when light can behave like a particle in its interaction with electrons, and it may be added that the tracks of the recoiling electrons have been photographed by the Wilson cloud chamber method.

177. THE EXISTENCE OF DISCRETE ENERGY STATES IN THE ATOM

Only one more link is needed to complete the chain of evidence concerning the corpuscular, or discontinuous, nature of radiation. It has been shown that radiation is corpuscular when it exchanges its energy with matter; it has been shown to be corpuscular when it preserves its existence; it only remains to show the existence of a corresponding discontinuity in matter itself. Since matter consists of atoms, the discontinuity must be sought in the atoms with which radiation interacts. This has been found in a number of ways and Fig. 271 is a diagram of one method used by Franck and Hertz in 1914. A metal filament F, coated with barium and strontium oxides, is heated with an electric current and emits a copious supply of electrons. They are accelerated to the grid G by a known potential V and most of them go on through the meshes of the grid to the plate P, which is at a slightly lower potential than the grid. The apparatus is enclosed in a glass vessel containing the element under test at such a pressure, that the electrons will suffer a number of collisions with the atoms of the gas between the filament and the grid, but, on the average, no collisions between the grid and the plate. This ensures that no electron shall get past the grid without having suffered at least one collision and, if it does get past the grid with enough velocity to reach the plate, it shall not be prevented from doing so by an encounter with an atom of the gas. When the filament current is switched on and the potential V is adjusted to a small value such as 1 volt and applied a current is registered on the galvanometer M. As the potential is raised the current increases steadily due to the increasing velocity

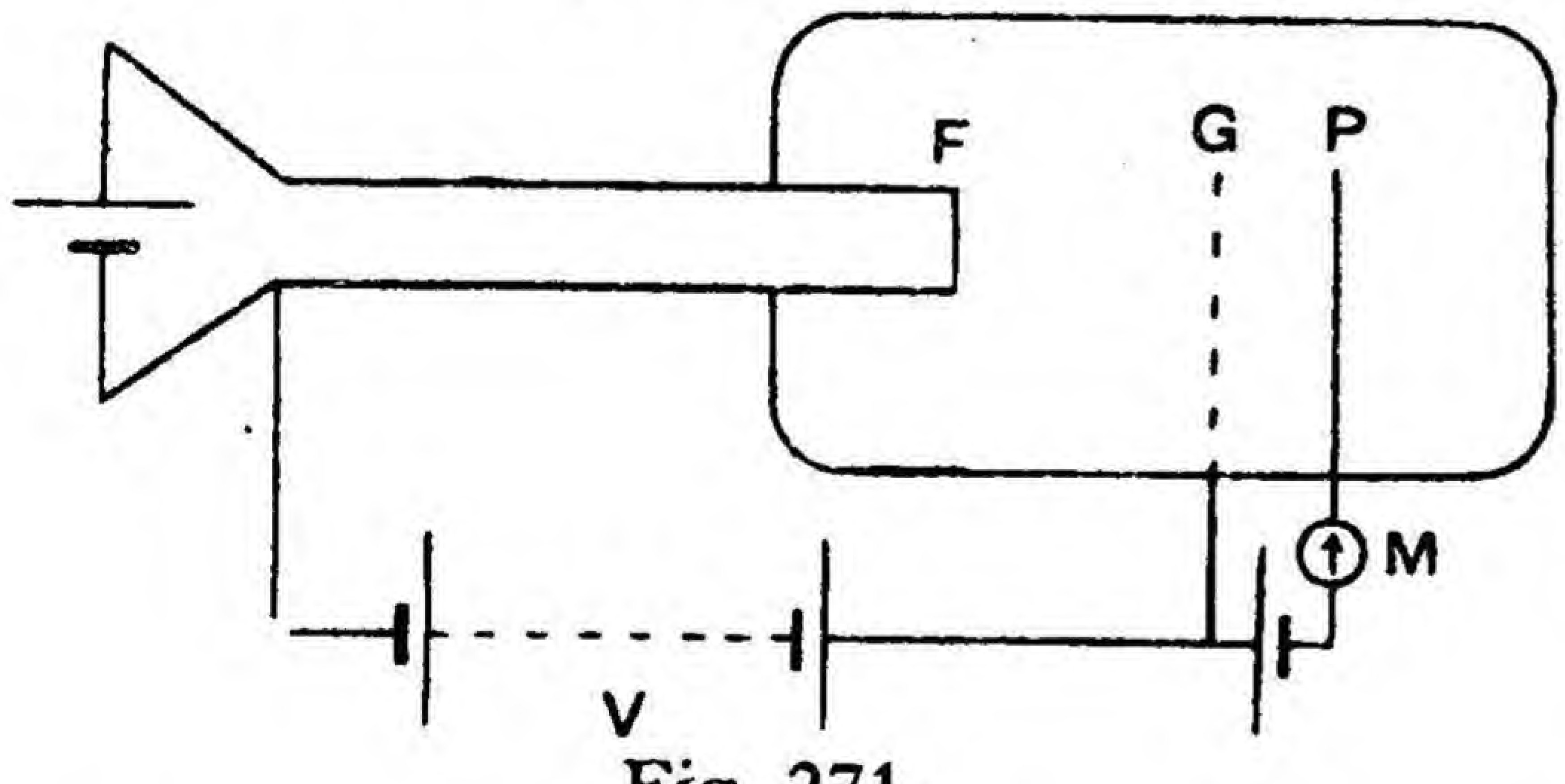


Fig. 271.

of the electrons, until a critical value V_1 is reached when it suddenly drops. A further increase in potential causes the current to increase once more. What is the meaning of this sudden drop? The explanation is as follows: while the potential V is less than V_1, the electrons make perfectly elastic collisions with the atoms and rebound with the same energy as they had before the impact; therefore their energy steadily increases as they travel along the electric field to the grid and, once they have passed through it, there is no collision to stop them reaching the plate P and passing on through the galvanometer M. But, when the critical potential V_1 is reached, the electron makes an **inelastic collision** with the atom, *giving up the whole of its kinetic energy to the atom to raise it from its normal state to another state of higher energy.* Hence the electron has now no energy left with which to overcome the small retarding potential between the grid and the plate, and therefore it cannot reach the plate; hence the current in the galvanometer drops to zero. So we have shown the existence of a state of the atom, in which its energy is greater by a finite amount than that in its normal state. We should expect the current to go on rising as the potential is increased beyond V_1 until it reaches $2V_1$, when a further sudden drop should occur owing to the electron having made its second inelastic collision by the time it reaches the grid. This and the corresponding drops at $3V_1$, $4V_1$, and so on have been detected.

Will the atom remain in this **excited state**, as the state of higher energy is called? Since systems tend to get to the condition of least energy, it is to be expected that the atom will return to the normal state, giving out this energy which it has received from the electron. In what form will it give up the energy? Radiation is quite a likely form and it is found that when V is greater than V_1, radiation of a definite wave-length is emitted. For example, with mercury vapour, a line of wave-length 2537×10^{-8} cm. or frequency $1{\cdot}18 \times 10^{15}$ c.p.s. is produced. Now the critical potential for mercury vapour is 4·90 volts, so that the energy given up when the atom returns to its normal state is the kinetic energy of an electron which has fallen through 4·90 volts, which is $4{\cdot}90 \times 10^8 \times 1{\cdot}60 \times 10^{-20}$ ergs, remembering that there are 10^8 e.m.u.'s. of P.D. in 1 volt and the charge of the electron is $1{\cdot}60 \times 10^{-20}$ e.m.u.'s. This energy is $7{\cdot}85 \times 10^{-12}$ ergs, and when it is divided by the frequency of the emitted radiation, the result is $6{\cdot}65 \times 10^{-27}$ ergs $\times$ secs., which is Planck's constant, h. Therefore we have proved, not only that the atom can exist in a set of discrete energy states but also that the emission of radiation is controlled by the familiar equation

$$E = h\nu \qquad \qquad (89)$$

where E is the amount of energy lost by the atom in changing from the excited to the normal state and ν is the frequency of the emitted radiation. We now see why any spectral line can be expressed as the difference of two terms (Art. 171); the terms from which a given set of spectral lines can be derived, correspond to a set of discrete energy states of the

atom emitting the lines and a given line is emitted, when the atom changes from one state to another of lower energy. The word discrete above means that there is a finite difference in energy between one state and the next. This property of the states is well brought out in Franck and Hertz's experiments, in which the electron gives *no* energy to the atom until it possesses a certain finite amount, when it transfers the *whole* of that energy to the atom.

This correlation of the terms with a set of discrete energy states of the atom has been verified by an extension of Franck and Hertz's experiments. An element in the form of a vapour is bombarded with electrons, which have fallen through a known potential, and the potentials at which a new line appears in the spectrum of the light emitted are measured. It is found that there are a set of values for a given element, and these potentials are a measure of the work needed to take the atom from its normal state, when its energy is least, to each of the other energy states in which it can exist. For hydrogen, the values of the potentials are 10·2, 12·0, 12·7, 13·0, and 13·5 volts respectively. The frequencies of any of the lines emitted, when hydrogen is excited by the above potentials, can be calculated in the same way as was done for the first excited state of mercury vapour, bearing in mind that, in the case of hydrogen, the atom may go from its excited state to the normal state via one of the intermediate states. The calculated frequencies agree with the observed frequencies in every case. Finally a potential is observed at which the complete spectrum of the hydrogen atom is produced, indicating that the electron has had sufficient energy to raise the atom to the highest energy state which it possesses. This state corresponds to the complete removal of one of the planetary electrons from the sphere of influence of the nucleus of the atom; in fact, the atom has been ionised. This potential is called the **ionisation potential** of the element; it is 13·5 volts for hydrogen, the last of the potentials given above. The relation between these critical potentials and the spectral terms is now quite clear; if we call the normal state of the atom its first state, the excited state next above that the second state, and so on, the frequency of the second term, say, is related to the difference in energy of the ionised state and the second state by equation (89), this relationship being illustrated in Fig. 272.

Our guess that the existence of a set of discontinuous terms in the spectra of atoms means discontinuity in the atom itself has been verified by direct experiment and has been correlated with the quantum views of radiation. **An atom can assume a set of energy states differing by a finite amount from each other; no other energies are permitted to it and the atom can only absorb or emit the quantity of energy needed to transform it between any two of these states.** So there can be no doubt that radiation has in some sense a corpuscular nature; for there is discontinuity in the behaviour of atoms themselves, there is a corpuscular nature in radiation when it remains as radiation,

and there is discontinuity when energy is transformed from radiation to atoms or from atoms to radiation. We are now faced with a serious problem presenting grave difficulties; how we are to reconcile these ideas with the wave theory of radiation, which is based on the undoubted

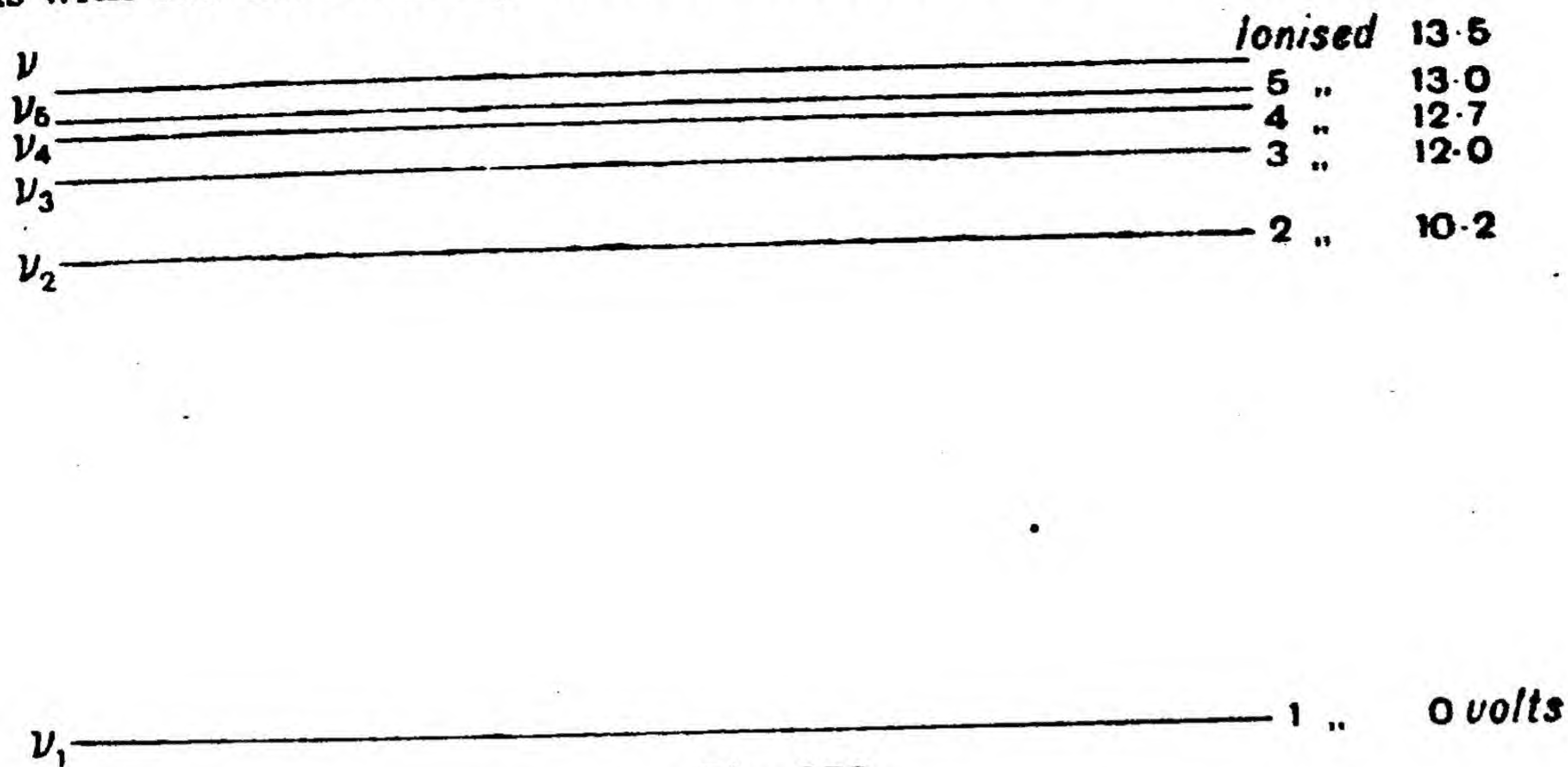


Fig. 272.

facts of interference and diffraction and which can never be explained on a purely corpuscular theory? Before attempting any answer to this question, we shall show that this problem is a fundamental one, striking at the very roots of physical science.

178. THE WAVES OF AN ELECTRON

We have already seen that the electron is the unit, or atom, of negative electricity, having a charge of $1{\cdot}60 \times 10^{-20}$ e.m.u.'s. and a mass of $9{\cdot}1 \times 10^{-28}$ gm., which is $\frac{1}{1838}$ that of the hydrogen atom. Its charge is so small that one billion electrons must pass each second through a good moving-coil galvanometer fitted with a lamp and scale in order to produce an appreciable deflection! It is possible to calculate its diameter from the above facts, making certain assumptions, and the value obtained is of the order of 10^{-13} cm. It is not surprising that the electron has always been treated as a particle and its properties have always agreed with that view. But some experiments were carried out by Professor G. P. Thomson in 1928 which showed that, under certain conditions, electrons behave as if they were guided by a group of waves. He produced a beam of electrons by applying a P.D. of some 60,000 volts to the terminals of a discharge tube and sent the electrons through a gold foil about 10^{-6} cm. thick. If the electrons behave as particles, it is to be expected that most of them will go straight on, a few being deflected by other electrons round the nuclei of the gold atoms, the greater the angle of deflection the less the number of electrons which will suffer that deflection. If a photographic plate is placed on the far side of the gold film we should expect a diffuse central spot due to the large number of undeflected electrons, growing

much feebler towards the outside. But the actual appearance, which is shown in Plate VI, reproduced by the kind permission of Professor G. P. Thomson, is quite different. It is a set of concentric rings of different radii ! The explanation of the rings is that the gold film consists of a number of tiny crystals arranged at random and each crystal is behaving as a diffraction grating for the electron waves, just as rock-salt behaves as a diffraction grating for X-rays (Art. 169). It will be remembered that the glancing angle of incidence has to assume certain values for strong reflection to occur, when the beam was deflected through twice the incident angle. If the permitted values for the electron waves for gold crystals are θ_1, θ_2; θ_3, and so on, reflection will occur for just those crystals which happen to make these angles with the incident beam of electrons, producing rings whose radii subtend angles $2\theta_1$, $2\theta_2$, $2\theta_3$, and so on at the film. When the constants of the gold crystal are known and the angular diameter of the rings have been measured, the wave-length of the electron waves can be calculated and is of the order of 10^{-8} cm. Precisely similar results were obtained by Davisson and Germer working at the Bell Telephone Laboratories, New York, using much slower electrons produced by accelerating them through a P.D. of the order of 100 volts. They reflected the electron beam from the surface of a single crystal of nickel, keeping the angle of incidence constant. They found that the reflected beam is strongest in a direction making the same angle with the surface of the crystal as the incident beam, and that the intensity varies with the velocity of the incident electrons, being a maximum for certain velocities. This suggests that the wave-length of the waves guiding the electrons depends on their velocity and that the velocities giving the maximum reflected intensities have wave-lengths which are copiously reflected at the given angle of incidence ; in fact, wave-lengths λ satisfying the Bragg equation

$$n\lambda = 2d \sin \theta$$

where θ is the glancing angle of incidence of the beam, n is an integer, and d is the grating element for nickel.

These experiments were inspired by the theory of de Broglie, who noticed that the laws of mechanics in their most general form are similar to Fermat's principle of stationary time, which expresses the laws of geometrical optics. He therefore put forward the idea that a particle takes the path which it does because it is being guided by waves satisfying the laws of geometrical optics. He showed that the wave-length of these waves is given by the relation

$$\lambda = \frac{h}{mv}$$

where m and v are the mass and velocity rexpectively of the particle and h is Planck's constant, which crops up again and is evidently of funda-

mental importance in these new ideas. The wave-lengths obtained from the experiments of Thomson and Davisson and Germer agree with the values calculated from the above relation. So the dual nature of radiation, which seemed so disconcerting, now finds a parallel in the electron, which is sometimes a particle and sometimes waves. The same duality has been found more recently in protons. It is evidently a fundamental feature of the whole physical universe, which consists of electrons, protons, and photons. It is interesting to note in connection with this brief reference to the waves of an electron that they have already found two practical applications; one is the investigation of the surface films of gas on metals, which are too thin to absorb X-rays, but produce an effect on electron waves; the other is the possibility of making an electron microscope, in which the focussing of the electron beam is performed by electric or magnetic fields. It may be possible to take a photograph of an atom with this microscope, since the wave-length of the waves can be made comparable with that of the atom by using fast enough electrons. Photographs of some very small organisms too small to be seen by the ordinary microscope have already been taken. Industry is at last realising the possibilities of science and nowadays there is only a small time lag between the discovery of new knowledge and its practical application.

179. CONCLUSION

We have now come to the end of our journey and it is natural to turn round to survey the ground which we have won. The theorems of Geometrical Optics were built up to explain the natural lenses and mirrors around us. The conception of the thin lens emerged and has been invaluable in the design of optical instruments, such as the camera, projection lantern, telescope, and microscope. The images produced by the simple varieties of these instruments were coloured at the edges and were never quite in focus, and the help of Geometrical Optics was sought again to probe the causes of these defects. Thus the theory of the various lens aberrations was developed and has helped the lens designer in producing the best possible lens for a given purpose; in practice, some defects must be diminished at the expense of others, for our theory told us that the perfect lens giving a perfectly focussed image of a large object with a lens of large aperture can never be produced; it is contrary to the laws of Nature. Newton's work on the nature of white light was inspired by the desire to improve telescopic images and to eliminate the colouring at the edges; it led ultimately not only to the solution of that problem but also to the discovery of spectra, which bid fair to revolutionise the very basis of physical science.

As these more practical problems were being solved, others of a more fundamental nature have also claimed our attention. What is light?

We have seen how the corpuscular theory seemed the correct answer at first on account of its simple explanation of rectilinear propagation, but it was slowly displaced by the wave theory because it could correlate a wider range of facts ; the phenomena of interference and diffraction, not rectilinear propagation, turned out to be the facts giving the key as to the nature of light. Finally the phenomenon of polarisation required light waves to be transverse, and this introduces the first problem, with which this book must conclude. What is the nature of light waves ? What is the condition which is propagated through empty space ? How does the propagation take place ?

The second problem is : what is the nature of the ether ? How can it transmit waves at the high velocity of light ? Is it a kind of rare fluid, like air only much less dense ? If so, how can it transmit *transverse* waves ?

Finally, how are we to reconcile the dual nature of both matter and radiation ? Radiation seems to behave as waves when we are dealing with large numbers of photons, and as particles when the interactions of single photons are concerned. Matter, on the other hand, seems to behave as particles when we deal with large amounts of it, and as waves when we deal with the individual particles, such as electrons. May it not be possible that the particles are the reality, whether they be photons or electrons, and that they are guided by waves, whose amplitude at a given point at a given instant measures the probability of the particles being at that point at that instant ? This would account for the interference fringes produced by Young's Slits, for example, since the places of zero amplitude of the waves would be places through which no photons pass, that is, they would be the positions of the dark fringes, just as they were before. We cannot pursue these ideas any further, but we deliberately end with these queries, since this illustrates the very essence of scientific investigation. When new fields of knowledge have been won they reveal fresh problems whose solution will open up, in their turn, wider fields of experience. The search for truth is eternal and as it goes on the underlying conceptions, which are used to correlate the experimental facts, become more and more abstract. We have seen the conception of light change from a stream of material corpuscles, which are easily imagined, into a wave travelling in the ether. The idea of a wave is not specially difficult to conceive, since there are concrete examples, such as water waves and sound waves. But the light wave is an imaginary wave of such simple properties that it can be represented by a simple equation, while a water wave near the sea shore can only be represented by an infinite series ! Also we do not know the nature of the condition which is propagated in a light wave and the nature of the medium in which that condition exists is very strange. But these abstract waves do correlate in a beautiful and satisfying way a wide range of facts, reducing to the corpuscular theory in the region for which that theory is true in the sense that the light rays travel

in straight lines. The wave theory triumphed in spite of the abstract nature of its fundamental conceptions, because it could explain all the facts which the corpuscular theory explained and, in addition, those facts of interference and diffraction for which that theory failed to account. Now the wave theory, in its present form, faces a similar crisis. It can never explain such facts as the photo-electric effect and the Compton effect; it is evident that we are on the threshold of a new and wider theory, which will reduce to the wave theory when applied to interference and diffraction, for which a wave theory will always be true, but taking the form of a quantum theory when applied to the photo-electric effect and Compton effect. Attempts at a synthesis of the two theories have been made in the Wave Mechanics and the reader must consult more advanced books for some account of this new hypothesis. But he will not have studied this book in vain if he approaches the new theory in the same way in which the corpuscular and wave theories have been discussed here; that is, in a critical attitude and guided by only two authorities, experimental fact and reason. If he thinks about the new ideas critically and humbly, it may be given to him to make some small addition to them, that they may explain the more perfectly the rich field of experience revealed by the masterly technique of modern experimental physics.

EXAMPLES ON CHAPTER XVII

1. Discuss the facts of the photo-electric effect and their explanation on both the wave-theory and the quantum theory.

2. If sunlight is absorbed by a body at the rate of 1 calorie per sec., how many quanta are absorbed per second? Take the wave-length of light as 6×10^{-5} cm. and h as $6{\cdot}55 \times 10^{-27}$ erg-secs.

3. What is the threshold frequency of a metal? The threshold wave-length for one of the alkali metals is $5{\cdot}45 \times 10^{-5}$ cm. Find the maximum velocity with which electrons are emitted if light of wave-length $4{\cdot}05 \times 10^{-5}$ cm. falls on the metal. Find also the positive potential in volts which the metal will acquire in order to stop any further emission. (Mass of the electron $= 8{\cdot}8 \times 10^{-28}$ gm., charge of the electron $= 4{\cdot}8 \times 10^{-10}$ E.S.U's., 1 E.S.U. of P.D. $= 300$ volts.)

4. X-rays of wave-length $1{\cdot}1 \times 10^{-8}$ cm. fall on a zinc plate. Find the maximum velocity of emission of the electron, if the threshold wave-length for zinc is $3{\cdot}0 \times 10^{-5}$ cm. Calculate also the energy of these electrons (a) in ergs, (b) in electron-volts. (The electron-volt is the energy acquired by an electron in falling through a P.D. of 1 volt.)

5. Discuss the Compton effect and its bearing on the nature of radiation.

6. Write an essay on the nature of light, bringing out the way in which the prevailing view has altered as more evidence has accumulated.

7. Do you consider that the accuracy of physical measurements is limited only by the coarseness of instruments? Do you think there is any absolute limit to the accuracy obtainable in measuring a length? *(Oxford Schol.)*

Answers

CHAPTER I

7. 37° 10′, 57° 55′.
8. 2°, 1·625.
9. 17·32 cm. below water surface, 13·1 cm. below liquid surface.

CHAPTER II

1. 15·3 cm.
3. 3·08 cm., 6·8 mm., 0·92 mm.
4. 14 cm., 29 cm., no.
6. Converging, 10 cm.
7. 40 cm., 120 cm.
8. 6·7 cm.
9. 1 cm., 22·5 cm.
10. 1·63, 23·05 cm., 25·13 cm.
11. 1·5 cm., 100 cm.
12. 1·20 cm., 18 cm.
13. 24·2 cm.
15. 13·8 cm., 1·575.
17. $u<f$ or $u>\frac{2d-f}{2(d-f)}$ where $d>f$.
18. 15·82 in., 6·33 in., 15·82 in., 7·91 in., 6·33 in., 3·16 in.
19. 6 in., 5·983 in., 3·590 in.
20. 1·856 in., 2·041 in., 31·8 in. × 15·9 in.
24. $\frac{50}{7}$ cm.
27. $12xy-70x-70y-125=0$, where x and y are measured from the centre of the combination.
28. 0·57 c.m., 8·6 c.m. beyond the second lens-image is further from second lens and longer.
29. $\pm\frac{\ }{2}$, ; $\frac{f}{2}$, $\frac{3}{2}$, f, 1 ; f, $\frac{1}{2}$.
31. $f_1{}^2u_2+f_2{}^2u_1=f_1{}^2f_2+f_1f_2{}^2$;
$f_1(u_2-f_2) \cdot f_1$.
$f_2(u_1-f_1) \cdot f_2$
33. 22·1 cm., 15·7 cm., 1·46.
34. −18·1 cm., −18·1 cm.
35. 30·0 cm. from lens, vertical lines, 0·5 mm. apart ; 120·0 cm. from lens, horizontal lines, 2 mm. apart.

CHAPTER III

1. 24·0 cm.
2. 39·4 in.
3. −20 in., 20 in.
4. Concave mirror of focal length 0·75 in.
5. 2 ft. behind mirror, 4·4 ft.
6. 5·14 ft.
8. 7·5 cm. towards the mirror.
9. 17·8 cm., 75 cm.
11. 40 cm.
12. −30·0 cm., 28·5 c.m., 31·8 cm., 1·50.
14. 114° 34′.
15. 5·2 cm., 9·4 cm.
17. 1·25.

CHAPTER IV

4. 2×10^{-10} sec.
6. 31·60 cm., 33·18 cm., 0·15 cm.

CHAPTER V

10. 0·18°.
11. 1·532, 40°.
12. 8°, 1·5, 1·6.
13. 1·516.
14. 46′, 16′.
15. 1·88 cm., 0·64 cm., 1·24 cm.
16. 2·4°, 0·023°.
20. 15 cm., −30 cm.
23. 1·655.
24. 50° 14′, 0·27°.
25. −36·2 cm. ; 20·14 cm., 19·78 cm. ; 44·75 cm., 44·70 cm., 44·75 cm.
26. 2·69 in., −4·86 in. ; crown glass convex lens, 4·96 in., −1·93 in. ; flint glass diverging lens, −1·93 in., −4·96 in.
27. 5·51 in., −9·99 in. ; 5·70 in., 13·0 in.
28. 2·60 in., −2·26 in., 7·08 in.

Answers

CHAPTER VII

1. 2·0 cm., 10 dioptres.
2. +50 cm., 50 cm., 2. Object at near point magnified 2×.
3. 50 cm., 21·39 cm., 0·14 cm., 1·43. **4.** −5·14 in.
5. 27·3 cm. to 300 cm. **6.** −85·7 cm.
7. −50 cm. for rays in a horizontal plane, −40 cm. for rays in a vertical plane; −6·0 dioptres spherical, 4·0 dioptres in a horizontal plane and 3·5 dioptres in a vertical plane.
8. 40 cm., −100 cm., 66·7 cm., 28·7 cm.
11. 5·94 in., 6·0 in., 7·13 in.; move lens 0·06 in. nearer to the plate, 53·5°.
12. 5·19 in.; crown glass lens, 2·52 in., 2·20 in., −3·30 in.; flint glass lens, −4·92 in., −3·20 in., ∞.

CHAPTER VIII

1. $\dfrac{5000}{5000-50d+d^2}$; 25 cm. from the eye.
5. 8·64 cm. beyond the second lens, 5·74 cm.
6. 10·1. **7.** 0·50°.
8. 45 cm. from the eye-piece.
10. 0·201 radians, 0·04 radians, 5·0; 5·0 ft.; move in the eye-piece 0·33 in., 6·0.
13. 4·26. **14.** 0·134 in., 1·68; 20·8 ft.
15. 4·12 in.
17. Crown glass lens, 2·24 ft., 2·07 ft., −2·64 ft.; flint glass lens, −4·06 ft., −2·64 ft., ∞.
18. 30 ft., 1·2 in., 3·8′; 36, 8⅓.
19. 3·2 cm., 11·2 cm. **22.** 5·35 cm. from the objective.

CHAPTER IX

2. 30 cm., 21·2 cm. **5.** 0·0035 foot-candles. **6.** 0·96.
8. 96 c-p., 2·83 ft. **10.** 3·5 cm., 55°. **11.** 15.
13. 100,000 lumens per sq. metre. **14.** 2860 candles per sq. ft., 0·74 ft.
15. 10,500 lumens per sq. ft., 330 candles per sq. ft.
16. 0·51 lux.
23. 0·98 ft.-candles, 29 candles per sq. ft.
25. 3 ft.-candles. **26.** 0·12.

CHAPTER XI

1. 7 days ±1·87 min. **3.** 221·4 rev. per sec. **6.** 0·875 mm., 2·02 mm.

CHAPTER XII

10. Increased to 61·9 cm.
14. 31·4 cm. per sec., 98·9 cm. per sec.^{2}, 27·2 cm. per sec., 8·66 cm.

CHAPTER XIII

3. 5880×10^{-8} cm. **7.** 0·169 cm. **10.** 1·13°.
17. 42·3 cm. **18.** 5460×10^{-8} cm. **19.** 1·590.
21. 0·366 cm. **22.** 0·0289 cm., 0·0578 cm.
27. 19. **29.** 1·000194.

CHAPTER XIV

7. 0·18 cm., 2 dark bands. **9.** 1442 cm.
11. 9538 lines. **12.** 3086 lines per cm.
13. 42′. **14.** 3010 lines per cm., 9° 55′.

15. 11′ 30″. **19.** 6 ; $6{\cdot}25 \times 10^{-5}$ cm., $5{\cdot}0 \times 10^{-5}$ cm., $4{\cdot}16 \times 10^{-5}$ cm.
20. 0·0315 cm.
22. $2e \sin \frac{2}{D} = n\lambda$, where D is the angle of minimum deviation ; $2{\cdot}59 \times 10^{-5}$ cm.
25. 17° 8′. **27.** 17° 8′, 17° 9′ ; yes.

CHAPTER XV

11. 0·00324 cm. **12.** 7·789 cm., 8·0°.

CHAPTER XVI

6. 6·14 cm., 1·370 cm. **7.** 35·42 cm., 30·00 cm., 24·15 cm.
8. 60·33 cm., 54·61 cm., 45·44 cm., 42·20 cm.
9. −46·29 km. per sec. **10.** −50 km. per sec.
13. −75 km. per sec. **14.** 0·0982° Å.U. no.
15. 6514·21 Å.U., 6557·79 Å.U. **17.** $1{\cdot}107 \times 10^{-8}$ cm., 23°17′, 36° 22′.

CHAPTER XVII

2. $1{\cdot}275 \times 10^{19}$ quanta. **3.** $5{\cdot}319 \times 10^{6}$ cm. per sec., 0·78 volts.
4. $6{\cdot}368 \times 10^{9}$ cm. per sec., $1{\cdot}78 \times 10^{-8}$ ergs, 11,150 electron volts.

Index

The numbers refer to pages